

**Azərbaycan Respublikası
Milli Təhlükəsizlik Nazirliyinin
Heydər Əliyev adına Akademiyası**

Vaqif Qasimov

**İnformasiya axtarışı üsulları
və sistemləri
(Dərslik)**

*Azərbaycan Respublikası Milli Təhlükəsizlik Nazirliyinin
Heydər Əliyev adına Akademiyasının Elmi Şurasının
03 dekabr 2014-cü il tarixli iclasında nəşrə tövsiyə olunmuş
və Təhsil Nazirliyinin 28 yanvar 2015-ci il tarixli
69 nömrəli əmri ilə ali məktəblər üçün
dərslik kimi təsdiq edilmişdir.*

Bakı-2015

Elmi redaktor: AMEA-nın həqiqi üzvü, t.e.d.,
prof. **Əli Abbasov**

Rəy verənlər: AMEA-nın həqiqi üzvü, f.-r.e.d.,
prof. **Fikrət Əliyev**

AMEA-nın müxbir üzvü, t.e.d.,
prof. **Ağası Məlikov**

t.e.d., prof. **Ələkbər Əliyev**

Qasimov V.Ə. İnformasiya axtarışı üsulları və sistemləri. Dərslük. Bakı: MTN-in Maddi-texniki Təminat Baş İdarəsinin Nəşriyyat-Poliqrafiya Mərkəzi. 2015, 288 s.

Texnika elmləri doktoru, professor Vaqif Qasimovun “İnformasiya axtarışı üsulları və sistemləri” dərsliyi internetdə informasiya axtarışı üsulları, vasitələri və sistemlərinin öyrənilməsinə həsr edilmişdir. Dərslükdə informasiya ilə bağlı əsas anlayışlar, informasiya tələbatı, onun formalaşması mərhələləri və öyrənilməsi üsulları, eləcə də internetin informasiya ehtiyatları və xidmətləri barədə geniş məlumat verilmişdir. Dərsliyə informasiya axtarışının dəqiqlik, dolğunluq, relevantlıq kimi əsas meyarları, modelləri, axtarışın effektivliyinin yüksəldilməsi üsulları, eləcə də internetin məşhur informasiya-axtarış sistemləri ilə bağlı mövzular daxil edilmişdir.

Dərslük Milli Təhlükəsizlik Nazirliyinin Heydər Əliyev adına Akademiyasında tədrisdə istifadə üçün nəzərdə tutulmuşdur. Dərslükdən respublikanın digər ali təhsil müəssisələrinin tələbə və müəllimləri, eləcə də bu sahədə çalışan elmi işçilər və mütəxəssislər də istifadə edə bilirlər.

© MTN-in Maddi-texniki Təminat Baş İdarəsinin
Nəşriyyat-Poliqrafiya Mərkəzi, 2015

MÜNDƏRİCAT

Giriş	8
I fəsil. İnformasiya və informasiya sistemləri	13
1.1. İnformasiya anlayışı, onun növləri	14
1.2. İnformasiyanın kodlaşdırılması və təsvir edilməsi	21
1.3. İnformasiyanın əsas xassələri	30
1.4. İnformasiya sistemləri	33
1.5. İnformasiya sistemlərinin təsnifatı	38
II fəsil. İnternetin informasiya ehtiyatları və xidmətləri	47
2.1. İnternet – qlobal informasiya şəbəkəsi	48
2.2. İnternet ünvanlar	52
2.3. İnternetin informasiya ehtiyatları	56
2.3.1. WWW – dünya hörümçək toru	57
2.3.2. E-mail – elektron poçtu	67
2.3.3. FTP – faylların ötürülməsi protokolu	70
2.3.4. Gopher – iyerarxik menyu	71
2.3.5. Telnet –uzaq məsafədən idarəetmə və sistemə daxilolma	72
2.3.6. BBS – elektron elan lövhələri	73
2.3.7. Usenet – telekonfrans sistemi	74
2.3.8. İnternetin interaktiv informasiya xidmətləri ...	75
2.4. İnternet şəbəkəsinin informasiya ehtiyatlarının xüsusiyyətləri	77

2.5. İnformasiya ehtiyatlarının məzmunlarının təhlilinə əsaslanan empirik paylanma qanunları	81
2.5.1. Zipfin I qanunu – “ranq-tezlik” asılılığı	82
2.5.2. Zipfin II qanunu – “say-tezlik” asılılığı	85
2.5.3. Bredford qanunu	87
2.5.4. Pareto qanunu	88
2.6. İnternetdə istifadə olunan əsas fayl formatları ...	89

III fəsil. İnformasiya tələbatı 93

3.1. İnformasiya tələbatı və onun məzmununa təsir edən amillər	94
3.2. İnformasiya tələbatının subyektləri	100
3.3. İnformasiya tələbatının öyrənilməsi və təqdim edilməsi üsulları	103
3.4. İnformasiya tələbatının modeli	104
3.5. İnformasiya tələbatının formalaşması mərhələləri	106
3.6. İnformasiya tələbatının təsnifatı	110
3.7. İnformasiya tələbatının tematik profilinin müəyyən edilməsi	112
3.8. İnformasiya ehtiyatının profilinin müəyyən edilməsi	115
3.9. Fərdin bir mövzu sahəsi üzrə informasiya tələbatının müəyyən edilməsi	116
3.10. Fərdin fəaliyyətinə aid olan bütün mövzu sahələri üzrə informasiya tələbatının müəyyən edilməsi	117
3.11. Bir mövzu sahəsi üzrə fəaliyyət göstərən bütün fərdlərin informasiya tələbatının məzmununun müəyyən edilməsi	118

3.12. Kollektiv informasiya tələbatının müəyyən edilməsi	119
--	-----

IV fəsil. İnformasiya axtarışının nəzəri əsasları ... 121

4.1. İnformasiya axtarışının əsas anlayışları	122
4.2. İnformasiya-axtarış dili	132
4.3. İnformasiya axtarışının nəticəsinin qiymətləndirilməsinin əsas meyarları	140

V fəsil. İnformasiya axtarışının effektivliyinin yüksəldilməsi üsulları 149

5.1. Axtarışın effektivliyinin qiymətləndirilməsinə yanaşmalar	150
5.2. Axtarışın dolğunluğunun yaxşılaşdırılması üsulları	151
5.3. Axtarışın dəqiqliyinin yaxşılaşdırılması üsulları	154
5.4. Relevantlığın yüksəldilməsi üçün istifadəçi ilə əks əlaqənin istifadəsi	156

VI fəsil. İnformasiya axtarışının modelləri 159

6.1. İnformasiya axtarışının modelləşdirilməsinə yanaşmalar	160
6.2. Nəzəri-çoxluqlar modeli	161
6.3. Vektorlar modeli	164
6.4. Statistik tezliklər modeli	166
6.5. Ehtimallar modeli	170
6.6. Linqvistik model	172
6.7. Semantik model	174
6.8. Qeyri-səlis model	178

VII fəsil. Qeyri-səlis model əsasında informasiya axtarışı üsulları	183
7.1. Sənədlərin profillərinə görə informasiya fəzasının tematik kataloqlara bölünməsi	184
7.2. Tematik profilin seçilməsinin dəqiqliyinin artırılması və tematik kataloqun keyfiyyətinin yaxşılaşdırılması üsulları	190
7.3. İstifadəçinin qeyri-səlis sorğusuna əsasən daha relevant sənədlərin axtarılması alqoritmi	194
7.4. Qeyri-səlis üstünlük münasibətləri əsasında relevant informasiyanın axtarılması	199
7.5. Qeyri-səlis üstünlük münasibətləri əsasında paylanmış informasiya axtarışının təşkili	205
VIII fəsil. İnternetdə informasiya axtarışı	211
8.1. İnternetdə informasiyanın axtarılması və əldə eilməsi üsulları	212
8.2. İnternetin informasiya-axtarış sistemləri	215
8.3. İnternetin informasiya-axtarış sistemlərinin təsnifatı	219
8.3.1. Tematik (mövzu) kataloqlar	221
8.3.2. İnternetin axtarış maşınları	222
8.3.3. Meta-axtarış sistemləri	226
8.4. İnternetin informasiya ehtiyatlarının indeksləşdirilməsi və təqdim olunması mexanizmləri	230
8.5. Axtarış spamları və onlara qarşı mübarizə	236

IX fəsil. İnternetin məşhur informasiya-axtarış sistemləri	241
9.1. İnternetin informasiya-axtarış sistemlərinin yaranma və inkişaf tarixi	242
9.2. Google axtarış sistemi haqqında ümumi məlumat	246
9.3. Google axtarış sisteminin inkişaf tarixi	249
9.4. Google sisteminin axtarış alqoritmləri	255
9.4.1. İnternet saytların indeksləşdirilməsi alqoritmi ...	256
9.4.2. Sıralama (nizamlama) alqoritmi	257
9.4.3. Biliklər qrafı (Knowledge Graph)	258
9.4.4. Google hummingbird (kolibri) alqoritmi	259
9.4.5. Google sisteminin web-axtarış spamları ilə mübarizəsi	259
9.4.6. Google Panda	261
9.4.7. Google Penguin (Pinqvin)	262
9.5. Google axtarış sisteminin əsas məhsulları	263
9.6. Google sisteminin axtarış imkanları	272
9.7. Google axtarış sisteminin interfeysi	274
9.8. Google-da axtarış sorğusunun tərtib edilməsi prinsipləri və şərtləri	278
Ədəbiyyat	283
İnternetin əsas axtarış saytları	287

GİRİŞ

İnformasiya axtarışı yüz illər boyu insanları ciddi düşündürən problemlərdən biri olmuşdur və hələ də sona qədər həll edilməmiş məsələ kimi qalmaqdadır. İnsanları əhatə edən informasiya ehtiyatlarının həcmi artdıqca, onların arasından lazım olan məlumatların axtarılıb tapılması daha da çətinləşir. Ona görə də informasiya axtarışı üçün daha təkmilləşdirilmiş və mükəmməl üsul və vasitələrin işlənilib hazırlanması istiqamətində daim ciddi işlər aparılır.

Lakin informasiya ehtiyatlarının həcmının astronomik sürətlə artması, onların xarakterinin və formasının kəskin dəyişməsi, informasiyanın saxlanılması və təqdim olunması üçün yeni texnologiyaların meydana gəlməsi səbəbindən, mövcud axtarış üsul və vasitələri insanların informasiya tələbatlarını ödəməkdə aciz qalır.

İnternet şəbəkəsinin geniş yayılması, onun informasiya xidmətlərinin sürətli inkişafı virtual informasiya fəzasının, o cümlədən internet saytların və informasiya sistemlərinin, virtual informasiya saxlanclarının, virtual kitabxanaların və s. informasiya ehtiyatlarının yaranmasına gətirib çıxarmışdır.

Hələ ötən əsrin 80-ci illərində, internet şəbəkəsinin FTP, Gopher, E-mail, Telnet kimi xidmətlərinin bazasında kifayət qədər böyük həcmdə informasiya ehtiyatlarının yarandığı, toplandığı, emal olunduğu və ötürüldüyü vaxtlarda axtarış vasitələrinin işlənilib hazırlanmasına və tətbiqinə böyük tələbat yaranmışdı. İnternetdə ilk informasiya axtarışı vasitələrinin çox əvvəllər yaranmasına baxmayaraq, onların həqiqi mənada vacibliyi yalnız web-texnologiyanın, yəni dünya hörümçək torunun meydana gəlməsindən sonra daha ciddi hiss olunmağa başladı.

Hazırda internetdə, onun nəhəng web-fəzasında saytların sayı və informasiyanın həcmi elə həddə çatmışdır ki, keçilməz kibercunqlini xatırladan internetin nəhəng informasiya fəzasında lazım olan məlumatın tapılması ciddi problemə çevrilmişdir.

Web-texnologiyanın bazasında duran hipermətn texnologiyası hiperkeçidlər vasitəsilə həmişə hara getmək lazım olduğunu göstərir. Lakin nəzərə almaq lazımdır ki, bu hiperkeçidlər heç də həmişə lazım olan yerə gətirib çıxarmır. Belə ki, istifadəçi hiperkeçidlər vasitəsilə tamamilə onu maraqlandırmayan səhifəyə düşə və ya müəyyən vaxtdan sonra əvvəl olduğu yerə – səhifəyə qayıda bilər.

Qeyd olunmalıdır ki, əsl cunqlidən fərqli olaraq, kibercunqlidə geriye yol həmişə yadda saxlanılır və istənilən vaxt addım-addım geri qayıtmaq mümkündür. Lakin kibercunqlidə ilkin yerə geri qayıtmaq üçün də böyük vaxt, maddi və şəbəkə resursları tələb olunur.

İnternetdə sənəd axınlarının gücünün astronomik sürətlə artması və yuxarıda göstərilən digər problemlər intellektual informasiya-axtarış vasitələrinin işlənilib hazırlanmasını və tətbiqini tələb edir.

Əlbəttə, bu problemin həllində web-texnologiya bazasında qurulmuş informasiya-axtarış sistemləri xüsusi rol oynayır. İnternetin informasiya-axtarış sistemləri – nəhəng informasiya fəzasında zəruri məlumatların axtarılıb tapılması üçün nəzərdə tutulmuş intellektual alətdir.

İndiyədək informasiya axtarışının modelləşdirilməsi, səmərəliliyinin və intellektuallığının artırılması istiqamətində xeyli elmi-tədqiqat və təcrübi işlərin görülməsinə baxmayaraq, bu sahədə hələ də arzu olunan nəticələr əldə edilməmişdir.

Bütün axtarış xidmətləri istifadəçilərin sorğularına cavab olaraq onların tələbatlarının ödənilməsi üçün daha münasib olan informasiya mənbələrinə istinadların siyahısını təqdim edir. Belə siyahıda lazımsız, istifadəçinin axtarış sorğusuna və tələbatına uyğun olmayan mənbələrə daha çox rast gəlinir.

Axtarış sorğusuna uyğunluq (relevantlıq) dərəcəsinin müəyyən edilməsi zamanı web-saytlarda reallaşdırılan naviqasiya qraflarının (hiperistinadların və hörümçək torunun) quruluşu, web-səhifələrdə aparılan dəyişikliklərin və onların həyat dövrünün (yaşama müddətinin) vaxtında nəzərə alınması böyük əhəmiyyət kəsb edir.

İnternetin informasiya-axtarış sistemlərində axtarışın həyata keçirilməsi və lazım olan məlumatın daha səmərəli yolla tapılıb istifadəçilərə təqdim edilməsi üçün əsas iki yanaşmadan istifadə olunur:

- mövzu (tematik) kataloqları;
- axtarış maşınları (indeksləri).

Hər iki yanaşma kifayət qədər inkişaf etmişdir və hazırda internetdə bu mexanizmləri reallaşdıran çoxlu sayda tematik kataloqlar və axtarış maşınları reallaşdırılmışdır.

Bu gün internetdə aşağıdakı məşhur informasiya-axtarış sistemləri müvəffəqiyyətlə fəaliyyət göstərirlər:

- **Google** (<http://www.google.com>,
<http://www.google.ru>,
<http://www.google.az> və s.);
- **Baidu** (<http://www.baidu.com>);
- **Yahoo** (<http://www.yahoo.com>);
- **Yandex** (<http://www.yandex.ru>);
- **Bing** (<http://www.bing.com>);
- **AltaVista** (<http://www.altavista.com>);
- **Lycos** (<http://www.lycos.com>);
- **HotBot** (<http://www.hotbot.com>);
- **Rambler** (<http://www.rambler.ru>);
- **Aport** (<http://www.aport.ru>);
- **STARS** (<http://www.stars.ru>);
- **Русский Интернет** (<http://www.rosit.ru>) və s.

İnternetdə işləyən hər bir istifadəçi axtarış nəticəsində ona təqdim olunan yüz minlərlə (hətta milyonlarla) sənəddə istinadların siyahısını gördükdə, özünə aşağıdakı kimi suallar verir:

- “axtarış sistemləri informasiya fəzasında sənədlərin axtarışını necə həyata keçirir?”;
- “axtarış sistemləri sənədlərin sorğuya uyğunluq dərəcəsini necə müəyyən edir?”;
- “axtarış sistemləri belə böyük və mürəkkəb siyahını necə formalaşdırır?”;
- “bu nəhəng siyahı ilə necə işləməli və onlar arasından lazım olan sənədləri necə seçməli?”.

Təqdim olunan vəsait yuxarıda sadalanan və sadalanmayan çoxlu sayda sualların cavablandırılması məqsədilə yazılmışdır. Belə ki, dərslük internetin informasiya massivində, yəni web-fəzada informasiya axtarışının təşkili və

həyata keçirilməsi mexanizmlərinin, eləcə də sənədlərin indeksləşdirilməsi və axtarılması üsul və vasitələrinin, informasiya-axtəriş sistemlərinin iş prinsiplərinin öyrənilməsinə həsr olunmuşdur.

Dərsləkdə informasiya axtəriş sahəsində mövcud olan əsas anlayışlar şərh edilir, istifadəçilərin informasiya tələbatları, onların formalaşması kriteriləri və mərhələləri, internetin informasiya ehtiyatlarının təsnifatına baxılır, informasiyanın indeksləşdirilməsi və axtəriş üzrə mövcud alqoritmlər, üsullar və modellər təhlil edilir, internet üçün nəzərdə tutulmuş informasiya-axtəriş sistemlərinin yaranma və inkişaf tarixi, arxitekturası və iş prinsipləri, eləcə də məşhur axtəriş sistemləri barədə məlumat verilir.

I FƏSİL

İNFORMASIYA VƏ İNFORMASIYA SİSTEMLƏRİ

İnformasiya anlayışı, onun növləri

İnformasiyanın kodlaşdırılması və təsvir edilməsi

İnformasiyanın əsas xassələri

İnformasiya sistemləri

İnformasiya sistemlərinin təsnifatı

1.1. İnformasiya anlayışı, onun növləri

Müasir dövrdə cəmiyyətin inkişafı bilavasitə informasiya ilə bağlıdır. Xarakterindən asılı olaraq insan fəaliyyətinin müxtəlif sahələrində elmi-texniki, iqtisadi, siyasi, idarəetmə, kommersiya və s. növ informasiya yaranır, toplanır, emal olunur və istifadə edilir. İnformasiya termini mənşəcə latın sözü olan «informatio» sözündən yaranmışdır, izahetmə, şərhətmə və məlumat vermə mənalarını daşıyır. İnformasiya informatikada ilkin anlayış kimi qəbul olunur.

İnformasiya anlayışına bütün tətbiq sahələrini əhadə edən dəqiq tərif vermək çətindir. Müxtəlif mənbələrdə informasiya anlayışının müxtəlif aspektlərdən aşağıdakı kimi izah və şərhələrinə rast gəlinir:

- informasiya – işarə və siqnallar vasitəsilə mövzu sahəsinin əks olunmasıdır;
- informasiya – xəbərlər, yeni faktlar, yeni biliklərdir;
- informasiya – məlumatın və ya xəbərin məzmunudur;
- informasiya – mövcud qeyri-müəyyənliyi aradan qaldıran məlumat və ya xəbərdir;
- informasiya – insanın ətraf aləmdə olan obyektlər və baş verən hadisələr haqqında məlumatlılıq səviyyəsini artıran xəbərlərdir;
- informasiya – fəlsəfi mənada real aləmin əksi kimi qəbul edilir.

Azərbaycan dilinin izahlı lüğətində informasiya anlayışı “*məlumatlar məcmusu*” kimi şərh olunur.

İnformatika elmi baxımından informasiyaya aşağıdakı tərif verilir.

İnformasiya – yaranma tarixindən, təqdimat formasından və təsnifatından asılı olmayaraq şəxslər, əşyalar, fakt-

lar, hadisələr və proseslər haqqında məlumatlar və bilgilərdir.

İnformasiya anlayışına Azərbaycan Respublikasının bir sıra qanunlarında tərif verilir.

- “İnformasiya, informasiyalaşdırma və informasiyanın mühafizəsi haqqında” Azərbaycan Respublikasının Qanunu (3 aprel 1998-ci il): “*İnformasiya – təqdimat formasından asılı olmayaraq şəxslər, əşya, fakt, hadisə və proseslər haqqında məlumatlardır*”.
- “İnformasiya azadlığı haqqında” Azərbaycan Respublikasının Qanunu (01 yanvar 2005-ci il): “*İnformasiya – yaranma tarixindən, təqdimat formasından və təsnifatından asılı olmayaraq səlahiyyətli orqanlar tərəfindən saxlanılan, maddi daşıyıcıda qeydə alınan faktlar, rəylər, bilgilər, xəbərlər və ya digər xarakterli məlumatların məcmusudur*”.
- “İnformasiya əldə etmək haqqında” Azərbaycan Respublikasının Qanunu (30 sentyabr 2005-ci il): “*İnformasiya – yaranma tarixindən, təqdimat formasından və təsnifatından asılı olmayaraq istənilən fəaliyyət nəticəsində yaradılan, yaxud əldə olunan faktlar, rəylər, bilgilər, xəbərlər və ya digər xarakterli məlumatlardır*”.

İnformasiyanı bir sıra əlamətlərə görə növlərə bölmək olar.

Qəbul edən reseptorların növündən asılı olaraq, qəbul etmə üsuluna görə informasiya aşağıdakı növlərə bölünür:

- ***vizual informasiya*** – görmə orqanları vasitəsilə qəbul edilən informasiya;
- ***audial informasiya*** – eşitmə orqanları vasitəsilə qəbul edilən informasiya;
- ***taktil informasiya*** – toxunmaqla duyan reseptorlar vasitəsilə qəbul edilən informasiya;

- *iyə bağlı informasiya* – iybilmə reseptorları vasitəsilə qəbul edilən informasiya;
- *dadla bağlı informasiya* – dadbilmə reseptorları vasitəsilə qəbul edilən informasiya.

Təqdim edilmə formasına görə informasiyanın növləri aşağıdakılardır:

- *mətni informasiya* – dilin leksemlərini işarə etmək üçün nəzərdə tutulmuş simvollar ardıcılığı şəklində ötürülən informasiya;
- *ədədi informasiya* – riyazi əməliyyatları işarə etmək üçün nəzərdə tutulmuş rəqəm və işarələr vasitəsilə verilən informasiya;
- *qrafik informasiya* – sxemlər, eskizlər, təsvirlər, qrafiklər, diaqramlar və simvollar şəklində təqdim olunan məlumat və ya verilənlər;
- *səs informasiyası* – dilin leksemlərindən istifadə etməklə şifahi danışmaq və ya audioyazı şəklində ötürülən informasiya;
- *video informasiya* – video qurğular vasitəsilə təsbit olunan, video yazılar şəklində saxlanılan və ötürülən informasiya.

Təyinatından asılı olaraq, informasiyanın aşağıdakı növlərini də fərqləndirirlər:

- *ictimai informasiya* – qanunlarla və ya digər normativ hüquqi aktlarla müəyyənləşdirilmiş ictimai vəzifələrin yerinə yetirilməsi prosesində yaradılan və ya əldə edilən faktlar, rəylər və bilgilərdir;
- *xüsusi informasiya* – xüsusiləşdirilmiş anlayışlar toplusundan ibarət olan, istifadə zamanı spesifik (kütlə tərəfindən anlaşılmayan, lakin müəyyən insanlar qrupu üçün başa düşülən və zəruri olan) məlumatları çatdıran informasiyadır.

- ***şəxsi və ailə həyatına dair informasiya (fərdi məlumat)*** – şəxsiyyəti birbaşa və ya dolayısı ilə identifikasiyaya imkan verən hadisələr, fəaliyyətlər, vəziyyətlər barədə faktlar, rəylər və bilgilərdir.

İstifadə sahəsindən və formasından asılı olaraq informasiyanı aşağıdakı növlərə bölürlər:

- ***sənədləşdirilmiş informasiya (sənəd)*** – maddi daşıyıcıda qeyd olunmuş və identifikasiyaya imkan verən olan rekvizitlərə malik informasiyadır;
- ***kütləvi informasiya*** – əldə olunması, istifadəsi, işlənməsi, başqa şəxslərə verilməsi və ya ötürülməsi Azərbaycan Respublikasının qanunvericiliyi ilə məhdudlaşdırılmayan, ümumi istifadə üçün nəzərdə tutulan sənədləşdirilmiş informasiyadır;
- ***konfidensial informasiya*** - əldə olunması, istifadəsi, işlənməsi, başqa şəxslərə verilməsi və ya ötürülməsi Azərbaycan Respublikasının qanunvericiliyi ilə məhdudlaşdırılan sənədləşdirilmiş informasiyadır.

Xarakterinə və formasına görə informasiyanın aşağıdakı növlərini qeyd etmək olar:

- ***tematik informasiya*** – texniki, texnoloji, marketing və digər mövzu sahələrinə aid olan informasiyadır;
- ***elmi-texniki informasiya*** – monoqrafiyalar, dissertasiyalar, avtoreferatlar, eləcə də elmi xarakterli məqalələr, referatlar, xülasələr, konfrans, telekonfrans və forum materialları və digər nəşrlərdir;
- ***reklam informasiyası*** – təşkilatlar (şirkətlər, firmalar və s.), onların rəhbərləri və əsas vəzifələri tutan şəxslər, təqdim olunan məhsul və xidmətlər barədə məlumatlar, müştəriləri maraqlandıran suallara cavablar və s.
- ***sorğu məlumatları*** – sorğu kitabçalarına analoji olan müxtəlif sorğu xarakterli məlumatlar, təşkilatların

ünvanları, web-saytlarına istinadlar, normativ bazalar və s.

- **xəbərlər** – baş verən hadisələr, onların mahiyyəti və inkişaf dinamikası barədə ilkin (işlənməmiş) məlumatlardır.

Gündəlik fəaliyyətdə informasiya anlayışı ilə yanaşı məlumat anlayışından da geniş istifadə olunur. Bəzən bu anlayışları eyniləşdirirlər. Lakin kontekstdən asılı olaraq onlar tamamilə fərqli mahiyyət daşıyırlar. Azərbaycan dilinin izahlı lüğətində məlumat anlayışı “*bir şey haqqında bilik, xəbərdarlıq*”, “*bir şey haqqında şifahi və ya yazılı arayış, bildiriş, izahat*”, “*xəbər, xəbərdarlıq*” kimi şərh olunur.

“İnformasiya azadlığı haqqında” Azərbaycan Respublikasının Qanununda məlumat aşağıdakı kimi şərh olunur: “**Məlumat** - *təbiətdə və cəmiyyətdə baş verən hadisələr, proseslər, təzahürər, fəaliyyətlər və əşyalar barədə faktlar, rəylər, bilgilərdir*”.

Beləliklə, demək olar ki, **məlumat** – ətraf mühitdə mövcud olan bütün maddi və mənəvi varlıqlar haqqında bilgilərdir, “**informasiya** isə məlumatın (xəbərin) məzmunudur”.

İnformasiya və məlumat anlayışları ilə yanaşı verilənlər və biliklər anlayışlarından istifadə olunur.

Verilənlər dedikdə konkret mövzu (fəaliyyət) sahəsində obyektlərin faktografik təqdimatını təmin edən konkret məlumatlar başa düşülür. Başqa sözlə, **verilənlər** – obyektlərin konkret göstəriciləri haqqında emal sisteminə daxil edilən faktiki məlumatlar toplusudur.

Bəzən aşağıdakı tərifə də rast gəlinir: “**verilənlər** – texniki və proqram vasitələrinin köməyi ilə emal olunması və rabitə xətti ilə ötürülməsi mümkün olan formallaşdırılmış şəkildə təqdim edilən faktlardır”.

“Elektron imza və elektron sənəd haqqında” Azərbaycan Respublikasının Qanununda “*verilənlər* – informasiya texnologiyaları vasitələri ilə emal edilməyə yararlı informasiya” kimi şərh olunur.

Ənənəvi olaraq, verilənləri bir neçə növə ayırırlar: ikilik (binar) verilənlər, mətn tipli verilənlər, ədəd tipli verilənlər və s. İkilik verilənlər yalnız onların strukturunu bilən xüsusi proqram təminatı vasitəsilə emal edilir, digər proqramlar isə onları dəyişdirmədən ötürür. Mətn tipli verilənlər informasiya sistemi tərəfindən hər hansı dildə yazılmış mətn kimi qəbul olunur. Adətən, bir sistemdən digərinə ötürülən zaman mətnlər kodlaşdırılır. Ədəd tipli verilənlər müəyyən kəmiyyətlərin qiymətlərinin verilməsi, hesablamaların aparılması üçün istifadə olunur.

Ümumiyyətlə, verilənlər obyekt haqqında və ya obyekt-dən daxil olan qeyd edilmiş məlumatlar çoxluğuudur. Verilənlər onların mümkün emalı proseslərinə adekvat olan konkret formalarda təqdim olunurlar. Informasiya kompüterə daxil edilən zaman o, verilənlərə çevrilir.

Bu baxımdan verilənləri faktoqrafik (faktiki) informasiya adlandırırlar. *Faktoqrafik informasiya* dedikdə ölçmə, təsviretmə və təsbitmə üsulu nəticəsində alınan təmiz (dəyişikliyə məruz qalmamış) məlumatlardır.

Verilənlərin informasiyaya çevrilməsi üçün onu aktuallaşdırmaq və mənaya malik məzmun vermək lazım gəlir. Ona görə də verilənlərin informasiya olması üçün onların qəbul və emal edilməsi üsulları və vasitələri olmalıdır. Başqa sözlə, informasiya verilənlər və onlar arasındakı münasibətlərdən yaranan dinamik məhsuldur.

Verilənlərin toplanması, saxlanması və emalı üçün verilənlər bazalarından istifadə olunur. *Verilənlər bazası* (VB) dedikdə hər hansı fiziki və ya virtual sistemin xa-

rakteristikalarını təsvir edən strukturlaşdırılmış, təşkil olunmuş verilənlər toplusudur. Verilənlər bazası anlayışını bəzən səhvən verilənlər bazalarını idarəetmə sistemləri (VBİS) ilə eyniləşdirirlər. Lakin yadda saxlamaq lazımdır ki, VB – bilavasitə verilənlərin toplusu, VBİS isə verilənlər bazalarının təşkili və aparılması üçün nəzərdə tutulmuş proqram sistemidir.

Bir çox hallarda rəsmi sənədlərdə verilənlər termini əvəzinə məlumat terminindən, verilənlər bazası termini əvəzinə isə məlumatlar bazası terminindən istifadə olunur.

Biliklər – həll edilən məsələ, təqdim olunan əşya, obyekt və mövzu sahəsi haqqında müəyyən məlumatlılıq səviyyəsinə uyğun olan və onun haqqında tam təsəvvür yaradan məlumatlar toplusudur. Azərbaycan dilinin izahlı lüğətində bilik “*oxumaq və ya təcrübə nəticəsində əldə edilən məlumat, elm*” kimi şərh olunur. Biliklər mövzu sahəsində müəyyən olunmuş məsələlərin həll edilməsinə imkan verən məlumatlar və onlar arasındakı münasibətləri tənzimləyən qanunauyğunluqlar əsasında formalaşır. İnformasiya sistemləri nəzəriyyəsində bilikləri məntiqi nəticə çıxarma prosesində istifadə olunan, formallaşdırılmış və strukturlaşdırılmış informasiya adlandırırlar.

Bilik – insanın dərk etmə fəaliyyətinin nəticələrinin mövcudolma və sistemləşdirmə formasıdır.

Bilik – obyektiv reallığın subyektiv obrazı, başqa sözlə daxili və xarici aləmin insanın şüurunda təqdimatlar, anlayışlar, mülahizələr, nəzəriyyələr formasında adekvat əksidir.

Süni intellekt, biliklər bazaları və ekspert sistemləri baxımından biliklər dedikdə ətraf aləm haqqında verilənlər, faktlar, xəbərlər və nəticə çıxarma qaydaları, o cümlədən obyektlərin xassələri, proses və hadisələrin qanuna-

uyğunluqları barədə informasiyanı, qərarların qəbul edilməsi üçün bu informasiyanın istifadə olunması qaydalarını özündə saxlayan toplu başa düşülür.

Biliklər təbii və süni dillər vasitəsilə təsbit olunur və saxlanılır. Təbiətinə görə biliklər deklarativ və prosedur bilikləri kimi iki kateqoriyaya bölünür. Deklarativ biliklər özündə yalnız müəyyən anlayışların strukturu haqqında məlumatları saxlayır. Onlar verilənlərə və ya faktlara daha yaxındır. Prosedur biliklər isə yeni biliklərin alınması, biliklərin yoxlanılması vasitələri və yolları haqqında məlumatları müəyyən edir.

Özündə bilikləri saxlayan verilənlər bazalarına biliklər bazaları qeyilir. **Biliklər bazası** (BB) – biliklərin idarə edilməsi, toplanması, saxlanılması, axtarılması və təqdim edilməsi üsul və vasitələrini reallaşdıran xüsusi növ verilənlər bazasıdır. Başqa sözlə, **biliklər bazası** dedikdə informasiyanın dərk edilmiş (şüurlu) emalına və məntiqi nəticə çıxarılmasına imkan verən faktlar (biliklər) və qaydalar toplusu başa düşülür.

Süni intellektin biliklər bazalarını və biliklərlə iş üsullarını öyrənən bölməsi **biliklər mühəndisliyi** adlanır.

1.2. **İnformasiyanın kodlaşdırılması və təsvir edilməsi**

Müasir dövrdə kompüter elmlərinin başlıca vəzifəsi informasiya texnologiyaları və sistemləri vasitəsilə informasiyanın qəbulu, emalı, saxlanması və ötürülməsindən ibarətdir. Bu baxımdan informatika elmində informasiya iki yerə bölünür: **analoq** və **rəqəmsal** informasiya.

Bu cür bölgünün aparılması informasiyanın təbiətindən irəli gəlir. Belə ki, təbiətdə və həyatda baş verən, insanın vizual müşahidə (görmə) və qulaqasma orqanları tərəfindən qəbul edilən, eləcə də klassik telefon, televiziya, radio və digər qurğular vasitəsilə ötürülən məlumatlar analoq informasiyadır.

Hesablama texnikasında, o cümlədən kompüterlərdə, kompüter şəbəkələrində və sistemlərində, müasir rabitə vasitələrində məlumatlar rəqəmsal informasiya şəklində yaradılır, saxlanılır, emal olunur və ötürülür.

Analoq informasiya ilə rəqəmsal informasiya arasında əlaqəni göstərmək üçün aşağıdakı nümunəyə baxaq. Ətrafda görünən təbiətin (məsələn, dağın, gölün) rəssam tərəfindən çəkilən və insanların görmə orqanları (gözləri) tərəfindən qəbul edilən şəkil, peyzaj analoq informasiyadır. Lakin həmin şəkilin (və ya təbiətin) rəqəmsal fotoaparata (müasir mobil telefonla) çəkilmiş variantı, fotoaparata, telefonun və ya kompüterin yaddaşına yazılmış rəqəmsal informasiya olur. Beləliklə, analoq informasiya rəqəmsal informasiyaya çevrilmiş olur.

Analoji olaraq, mətn, şəkil, audio və video məlumatlar da kompüterə daxil edilərkən kodlaşdırılır və siqnallara çevrilir, yəni rəqəmsal informasiyaya çevrilmiş olur.

Analoq informasiya kəsilməz olur, yəni bu informasiyada hadisənin bütün məqamları öz əksini tapır, lakin rəqəmsal informasiya baş vermiş hadisənin reallığını itirməmək şərti ilə onu ayrı-ayrı kadrılar, yəni siqnallar şəklində saxlayır. Bu baxımdan, rəqəmsal informasiyaya diskret informasiya deyirlər. Analıq informasiyanın rəqəmsal informasiyaya çevrilməsi prosesinə analıq-rəqəm çevrilməsi deyilir.

İnsanın hissiyyat orqanları yalnız analıq informasiyanı qəbul və emal edə, saxlaya və ötürə bilər. Qeyd olunduğu

kimi, analoq informasiyanın qəbulu, ötürülməsi və emalı funksiyalarını yerinə yetirən bir çox texniki analoq qurğuları və vasitələri (məsələn: televizor, radio, maqnitafon, telefon) mövcuddur. Lakin müasir dövrdə analoq qurğular yeni nəsil rəqəmsal qurğularla əvəz olunur. Hazırda rəqəmsal informasiya ilə işləyən qurğulara misal olaraq kompüter texnikasını, rəqəmsal televizorları, rəqəmsal raddionu, DVD və BluRay qurğularını, rəqəmsal telefonları və s. misal göstərmək olar. Belə qurğular yalnız rəqəmli informasiyanın qəbulu, emalı, saxlanması və ötürülməsi üçün nəzərdə tutulmuşdur.

Qeyd olunduğu kimi, informasiya siqnallar formasında qəbul olunur, emal edilir və ötürülür. Hər növ siqnalın öz qəbuledicisi olmalıdır. Məsələn, qulaq səs siqnallarını, göz işıq siqnallarını və s. qəbul edir. İnformasiyanın saxlanması üçün isə onun kodlaşdırılması tələb olunur. İnsan düşüncəsinin və fikirlərinin mətnlər şəklində yazılmasını informasiyanın xüsusi siqnallarla kodlaşdırılması kimi qəbul etmək olar. İnformasiyanın yazı şəklində saxlanması üçün qəbul edilmiş siqnallar hərflərdir. Yazının kodlaşdırılması sisteminə əlifba deyilir. Əlifbalar müxtəlif olduğundan eyni bir informasiyanı müxtəlif sistemlərdə (dillərdə) kodlaşdırmaq mümkündür. Bəzən yazıların kodlaşdırılması üçün daha mürəkkəb kodlaşdırma sistemlərindən – heroqliflərdən istifadə edirlər.

Rəqəmsal informasiyanın elektron qurğuların, o cümlədən kompüterlərin köməyi ilə saxlanması, emalı, qəbulu və ötürülməsi üçün onlar da kodlaşdırılmalıdır. Kompüterdə məlumatlar 0 və 1 rəqəmlərinin ardıcılığı şəklində ifadə edilir, yəni rəqəmsal informasiyanın verilməsi (təqdim və təsvir olunması) üçün istifadə olunan siqnallar 0 və 1 rəqəmləri ilə işarə olunur. Məsələn, qurğulardan asılı

olaraq impulsun, işıq selinin, elektrik cərəyanının və s. olması 1, olmaması isə 0 rəqəmi ilə verilir.

Kompüterdə informasiya ayrı-ayrı simvolların toplusu şəklində emal edilir. İnformasiyanın verilməsi üçün tələb olunan bütün simvollar müəyyən olunmuş kodlarla – ədədlərlə işarə olunur. Hər hansı söz kompüterə daxil edilərkən və ya rabitə xətti ilə ötürülərkən o, ayrı-ayrı hərflərinin (simvollarının) kodlarına, yəni ədədlərə çevrilir və ədədlər şəklində saxlanılır və ya ötürülür. Kompüterlərdə, eləcə də digər rəqəmsal texnologiyalarda audio, video, qrafika, şəkil və s. növ məlumatlar da analoji qaydada kodlaşdırılır, emal olunur və saxlanılır. Bu kodlar da (ədədlər) iki rəqəmin – 0 və 1 rəqəmlərinin vasitəsilə ifadə olunur.

İnformasiyanın 0 və 1 rəqəmlərinin vasitəsilə ifadə olunması sistemi *ikilik say sistemi* adlanır. İkilik say sistemi ingiliscə "*binary digit*" adlanır.

İnformasiya texnologiyalarının xüsusiyyətlərindən irəli gələrək, informasiyanın həcmnin ölçülməsi üçün *bit* və *bayt* baza vahidlərindən istifadə edilir. Digər ölçü vahidləri baytdan törəmə ölçü vahidləridir.

Bit (ingiliscə *binary digit* sözlərindən) – informasiyanın saxlanması üçün istifadə olunan ən kiçik ölçü vahidi olub yaddaş qurğularının fiziki texnologiyasından irəli gəlmiş və informasiyanın saxlanması üçün nəzərdə tutulmuş elementlərlə müəyyən olunmuşdur. Məlum olduğu kimi, 1 bit bir ikilik simvola, yəni 0 və ya 1 rəqəmlərinə uyğun gəlir.

İnformasiyanın emalı zamanı bitlərin istifadəsi əlverişli olmur. Bu məqsədlə baytlardan istifadə olunur. *Bayt* – verilənlərin və ya kompüterin yaddaşının ünvanlarına (müraciət oluna) bilən ən kiçik vahididir. 1 bayt 8 bitdən ibarətdir. İlk baxışdan elə yanlış təsəvvür yarana bilər ki, baytda saxlanılan informasiyanın miqdarı onun hər hansı

bitindəki informasiyanın həcmindən 8 dəfə çoxdur. Əslində baytda informasiyanın həcmi üçün bitlərin sayı deyil, eyni zamanda onların sıradakı mövqeyi də əhəmiyyət kəsb edir. Belə ki, 00000001, 00001000, 10000000 baytları bitlərin, yəni 1 və 0 rəqəmlərinin sayının eyni olmasına baxmayaraq, müxtəlif ölçülü məlumatları (uyğun olaraq, 1, 8 və 128 ədədlərini) ifadə edirlər.

Onluq say sistemində olan ədədlərin ikilik say sistemində ekvivalentinin tapılması, ədədlərin bir sistemdən digərinə çevrilməsi qaydaları ilə müəyyən edilir. Aydındır ki, 1 baytın, yəni 8 bitin köməyiylə 256 (0-dan 255-ə qədər) ədədi ifadə etmək mümkündür:

- | | | | |
|----|-------------|------|---------------|
| 1. | 0000 0000=0 | 8. | 0000 0111=7 |
| 2. | 0000 0001=1 | 9. | 0000 1000=9 |
| 3. | 0000 0010=2 | ... | |
| 4. | 0000 0011=3 | 253. | 1111 1100=252 |
| 5. | 0000 0100=4 | 254. | 1111 1101=253 |
| 6. | 0000 0101=5 | 255. | 1111 1110=254 |
| 7. | 0000 0110=6 | 256. | 1111 1111=255 |

Ümumiyyətlə, praktikada aşağıdakı informasiya ölçü vahidlərindən istifadə olunur:

Bit – bir ikilik rəqəmdən (0 və ya 1 rəqəmindən) ibarət ən kiçik ölçü vahidi olub impulsun daxil olmasını (1) və ya olmamasını (0) bildirir.

Bayt – (**b** kimi işarə olunur) kompüterdə istifadə edilən simvolların (rəqəmlərin, hərflərin və digər işarələrin) ikilik say sistemində təsvir edilməsi üçün istifadə olunan və 8 bitdən ibarət olan ölçü vahididir:

$$1 \text{ bayt} = 8 \text{ bit.}$$

Məsələn, 210 ədədi ikilik say sistemdə "11010010" kimi 1 bayt şəklində ifadə olunur ki, bu da 8 ikilik rəqəmdən, yəni 8 bitdən ibarətdir.

Kilobayt – (**Kb** kimi işarə olunur) adından göründüyü kimi ("kilo" sözü burada "ikilik min"i bildirir) təxminən min bayt ölçüsünə uyğun gəlir. Lakin ikilik say sisteminin xüsusiyyətlərdən asılı olaraq, kilo 1000-ə deyil, 1024-ə (1 kilo = 2^{10}) bərabərdir. Beləliklə,

$$1 \text{ Kb} = 2^{10} \text{ b} = 1024 \text{ b.}$$

Meqabayt – (**Mb** kimi işarə olunur) aşağıdakı kimi qiymətləndirilir:

$$1 \text{ Mb} = 2^{10} \text{ kb} = 2^{20} \text{ b} = 1048576 \text{ b.}$$

Qiqabayt – (**Gb** kimi işarə olunur) vahidinin qiyməti belədir:

$$1 \text{ Gb} = 2^{10} \text{ Mb} = 2^{30} \text{ b} = 1073741824 \text{ b.}$$

Terrabayt – (**Tb** kimi işarə olunur) aşağıdakı qiymətə malikdir:

$$1 \text{ Tb} = 2^{10} \text{ Gb} = 2^{40} \text{ b.}$$

Bod – rabitə xətti vasitəsilə ötürülən zaman informasiya bodlarla ölçülür və 1 saniyədə ötürülən bitlərin miqdarını göstərir:

$$1 \text{ bod} = 1 \text{ bit/san.}$$

Aydındır ki, 1 bit vasitəsilə yalnız iki qiymət – 1 və ya 0 qiymətləri kodlaşdırıla bilər. 2 bitlə dörd (0-dan 3-ə qədər - 00, 01, 10, 11), 3 bitlə səkkiz müxtəlif qiymət (0-dan 7-yə kimi) kodlaşdırılır. Ona görə də, kodlaşdırma üçün istifadə olunan bitlərin sayı artdıqca, daha böyük həcmdə informasiyanın kodlaşdırılması mümkün olur.

Ümumiyyətlə, müasir kompüterlərdə 256 simvoldan istifadə edilir. Yuxarıda qeyd olunduğu kimi, bu 256 simvolun təsvir olunması üçün maksimum 8 bit kifayət edir. Onlardan 128 simvol standart, 128 simvol isə dəyişən olur. Standart hissəyə idarəedici simvollar, rəqəmlər, hesab əməlləri, böyük və kiçik ingilis hərfləri, durğu işarələri daxildir. Dəyişən hissədə isə milli şriftlər (milli standartlar), mətnlərdə istifadə olunan müxtəlif işarələr, psevdoqrafika və s. əksini tapır.

Qeyd olunduğu kimi, kompüterlər yalnız rəqəmlərlə işləyə bilir. Ona görə də kompüterlərin yaddaşında hərfləri saxlamaq üçün hər bir simvola, o cümlədən hərflər və ya rəqəmə müəyyən bir ədəd (kod) qarşı qoyulur. Əvvəllər simvolların kodlaşdırılması üçün müxtəlif simvollar çoxluğundan (standartlarından) istifadə olunurdu. ASCII, DOS, ISO, KOI, Windows və s. kodlarını belə standartlara misal olaraq göstərmək olar. ASCII standartı kompüterlərdə istifadə olunan ilk simvollar toplularından biridir. ASCII standartı kompüterlər üçün "əlifba" hesab olunur. Burada kompüterdə istifadə olunan simvollar, onların şəkilləri, kodları (nömrələri) müəyyən edilir.

Lakin bu standartlardan heç biri kompüter sistemlərində və digər hesablama texnikasında istifadə olunan bütün simvolları təsvir etməyə imkan vermir. Bu baxımdan müxtəlif sistemlərdə (Windows, Unix, Macintosh və s.) ayrıca bir dil üçün müxtəlif kodlaşdırmalardan istifadə olunurdu. Elə bir standart kodlaşdırma sistemi yox idi ki, o, bütün hərfləri, rəqəmləri, simvolları, işarələri və texniki simvolları özündə cəmləşdirsin. Eyni zamanda bir neçə dildə işləmək problem yaradırdı.

Bundan əlavə, çox vaxt bu kodlaşdırma standartları arasında uyğunluq olmurdu. Məsələn, mövcud iki müxtəlif

kodlaşdırma sistemində müxtəlif iki simvola eyni bir kod və ya eyni simvola iki müxtəlif kod verilir. Bu da kompüterlərdə, əsasən də server kompüterlərində müxtəlif kodlaşdırma sistemlərinin saxlanması tələb edirdi. Bu halda informasiyanın bir platformadan digər platformaya ötürülməsi və ya digər kodlaşdırmaya çevrilməsi müəyyən çətinliklər törədirdi.

Məhz bu baxımdan Unicode standartının qəbul olunması qeyd olunan problemləri həll etdi. Unicode ISO/IEC 10646 standartının rəsmi reallaşdırılmasıdır. Unicode standartı istifadə olunan dildən, kompüterlərin platformasından, proqram təminatından asılı olmayaraq bütün simvolara unikal kod mənimsədir. Unicode Apple, HP, IBM, JustSystem, Microsoft, oracle, SAP, Sun, Sybase, Unisys və s. kimi böyük kompüter şirkətləri tərəfindən kodlaşdırma standartı kimi qəbul edilmişdir.

XML, Java, ECMAScript (JavaScript), LDAP, CORBA 3.0, WML və digər müasir texnologiyalar tərəfindən yalnız Unicode standartı istifadə olunur. Eyni zamanda qeyd olunmalıdır ki, bu standart bütün əməliyyat sistemləri (platformaları), internet proqramları və müxtəlif proqram təminatları tərəfindən dəstəklənir. Unicode standartının dəstəklənməsi və istifadəsinin təmin edilməsi müasir proqram təminatlarının inkişafının ən vacib aspektlərindən biridir.

Müştəri-server texnologiyasında olan şəbəkə proqramlarında və əlavələrində, eləcə də internet şəbəkəsində istifadə olunan proqram sistemlərində informasiyanın təqdim edilməsi üçün Unicode standartının tətbiqi proqram məhsullarının hazırlanmasına və dəstəklənməsinə çəkilən xərclərin azaldılmasına gətirib çıxarır. Belə ki, Unicode müxtəlif platformalarda istifadə oluna biləcək vahid proqram təminatını və ya web-səhifəni hazırlamağa imkan verir.

Beləliklə, Unicode beynəlxalq standartlara uyğun olmaq şərti ilə bütün dünya ölkələri arasında elektron mətnlərin problemsiz mübadiləsinə və oxunuşunu dəstəkləmək üçün simvolların kodlaşdırılmış bir sistemidir. Bundan əlavə, o bütün dillərin klassik və tarixi mətnlərini dəstəkləyir.

Unicode standartı onun elektron variantı olan "Unicode Standart Annexes" (Unicode standart əlavələri) və "Unicode Character Database" (Unicode simvollarının verilənlər bazası), həmçinin "Standart Unicode" kitabı vasitəsi ilə təyin olunmuşdur. Hər bir versiyanın dəqiq tərkibini müəyyənləşdirmək üçün «Enumerated Versions of the Unicode Standart» (Unicode standartının nömrələnmiş versiyaları) standartından istifadə olunur. Unicode 4.0 versiyası Unicode standartının son variantıdır.

Azərbaycan əlifbası daxil edilmiş standart Unicode şriftləri aşağıdakılardır:

- Arial;
- Times New Roman;
- Courier;
- Verdana;
- Lucida Sans Unicode;
- Tahoma;
- Palatino Linotype və s.

Unicode standartı bütün müasir və qədim yazılı dillərin işarələrinin vahid sistem daxilində kodlaşdırılmasını təmin etmək məqsədi ilə işlənilmişdir. Standartda hər bir simvol 16 bit (2 bayt) vasitəsilə ilə təqdim olunur və 8 bitlik kodlaşdırmadan fərqli olaraq daha çox simvolun təsvir olunmasına imkan verir. Belə ki, Unicode standartı 65536 kodu əhatə edə bilər.

Unicode standartına hərf, rəqəm və vurğu işarələrindən ibarət simvollar qrupu, riyazi və texniki işarələr qrupu,

göstəricilər və s. daxildir. Burada "tilda" və bu kimi işarələrin müxtəlif variantları, müxtəlif işarələrin birləşməsindən əmələ gələn diakritlərin kodlaşdırılmasını təmin etmək olur. Unicode dünya əlifbalarının işarələri, heroqliflər və işarələrdən ibarət olan 29000-dən çox simvolların kodlarını özündə birləşdirir.

Unicode standartında ehtiyat kimi saxlanmış və daha 29000-dən çox simvolun istifadəsi üçün nəzərdə tutulmuş kodlar mövcuddur. Onların köməyi ilə gələcəkdə hətta tarixi heroqliflərin və ya mövcud əlifbaların mümkün genişlənmələrini nəzərə almağa imkan verir. Bundan əlavə, 6000-dən çox kod fərdi istifadə üçün nəzərdə tutulmuşdur ki, bunlar proqram təminatı və avadanlıq istehsalçıları tərəfindən xüsusi məqsədlər üçün istifadə oluna bilər.

Unicode standartının kodlarının ümumi diapazonu bir neçə altçoxluğa bölünmüşdür. Bu altçoxluqlar hər hansı dilin əlifbasından və ya funksiyalarından asılı olaraq xüsusi işarələr qruplarından ibarətdir. Unicode standartı haqqında daha ətraflı məlumatı <http://www.unicode.org/> ünvanından almaq olar.

1.3. İnformasiyanın əsas xassələri

İstənilən obyekt kimi informasiya da müəyyən xassələrə malikdir. İnformasiyanın xassələrinə həm verilənlərin, həm də informasiya prosesində verilənlərlə işləyən üsul və vasitələrin xassələri təsir edir. ***İnformasiyanın keyfiyyəti*** dedikdə onun istehlakçıların tələbatlarına uyğun olma dərəcəsi başa düşülür. İnformasiyanın xassələri onun istehlakçısının tələbatından asılı olub, nisbi xarakter daşıyır.

İnformasiyanın keyfiyyətini müəyyən edən *əsas xassələri* aşağıdakılardır.

Obyektivlik – informasiyanın özündə subyektiv fikirləri saxlamaması xassəsidir. Subyektiv fikir (qiymət) dedikdə baş verən prosesə, vəziyyətə və hadisəyə hər hansı insanın baxışı, fikri, mülahizəsi və ya rəyi başa düşülür. Təbii ki, eyni bir məsələ ilə bağlı baxışlar və fikirlər müxtəlif ola bilər. Subyektiv baxışdan, fikirdən, mülahizədən asılı olmayan informasiya obyektiv informasiya adlanır. İnformasiyanın obyektivliyi anlayışı nisbidir, çünki onun emalı üsulları, adətən, subyektiv olur. Ona görə də istənilən informasiya prosesinin gedişində informasiyanın obyektivlik dərəcəsi həmişə aşağı düşür. İnformasiya proseslərinin ən az subyektivlik daxil etdiyi informasiya daha obyektiv adlanır.

Dolğunluq (tamlıq) – informasiyanın əks olunan obyektə və ya prosesi tam xarakterizə etməyə imkan verən xassəsidir. Əgər informasiya həll olunan məsələnin mahiyyətinin başa düşülməsi və qərarın qəbul edilməsi üçün kifayət qədərdirsə, onu dolğun (tam) adlandırmaq olar. Tam olmayan informasiya səhv nəticənin çıxarılmasına və düzgün olmayan qərarın qəbul edilməsinə səbəb ola bilər. Lakin müəyyən vəziyyətlərdə informasiyanın izafiliyi düzgün qərarın qəbul edilməsinə əngəl yarada bilər.

Doğruluq – informasiyanın gizli səhvlərə malik olmaması xassəsidir. Başqa sözlə, doğruluq alınmış informasiyanın onun həqiqi mənasına uyğunluq dərəcəsi kimi başa düşülür. Əgər informasiya real (əsil) vəziyyəti təhrif etmirsə və mövcud obyektləri zəruri dəqiqliklə əks etdirirsə, onda deyirlər ki, informasiya doğrudur. Obyektiv informasiya həmişə doğrudur, lakin doğru informasiya həm obyektiv, həm də subyektiv ola bilər.

Doğru informasiya düzgün qərar qəbul etməyə kömək edir. İnformasiyanın məzmununun qəsdən təhrif edilməsi (dezinformasiya), subyektiv signalın təsadüfən, küylərin təsiri nəticəsində təhrif olunması, informasiyanın təsbit edilməsində səhvlərin olması və digər səbəblərdən onun doğruluğu pozula bilər. Ümumi halda, haqqında məlumat ötürülən hadisənin başvermə vaxtının göstərilməsi, müxtəlif mənbələrdən alınmış məlumatların tutuşdurulması, dezinformasiyanın vaxtında aşkarlanması, təhrif olunmuş informasiyanın istisna edilməsi yolu ilə informasiyanın doğruluğunu təmin etmək olar.

Adekvatlıq (dəqiqlik) – informasiyanın əks olunan obyekt, proses və ya hadisəyə birqiyətli uyğunluğu xassəsidir. Təqdim etdiyi obyektə həqiqətən uyğun olan informasiya adekvat informasiya adlanır. Əgər informasiyanın köməyi ilə real obyekt, proses, hadisə haqqında obyektiv təqdimat yaratmaq mümkündürsə, onda həmin informasiya adekvatdır.

Əlyetərlilik – informasiyanın istifadəçi tərəfindən əldə oluna bilməsi imkanları ilə xarakterizə olunan xassəsidir. Əlyetərlilik – informasiyanın əldə olunmasının mümkünliyünün ölçüsüdür. Əgər informasiyanın əldə edilməsi və çevrilməsi imkanları varsa, onda deyirlər ki, informasiya əl çatandır (əl yetəndir). İnformasiyanın əlyetərliliyi həm də qəbuletmə üçün onun təqdim edilməsi forması ilə müəyyən olunur. Əgər informasiya başa düşülməyən və ya qavrama üçün aydın olmayan şəkildə ifadə olunubsa, onda o, əldə olunmayan hesab edilir.

Aktuallıq (vaxtında olma) – informasiyanın cari vaxta uyğunluq dərəcəsini göstərən xassəsidir. Aktuallıq – hazırkı vaxt üçün informasiyanın vacibliyini və əhəmiyyətliliyini müəyyən edir. Aktuallıq termini əvəzinə bəzən “vax-

tında olma” (təzəlik) termini işlədirlər. Vaxtında olma dedikdə informasiyanın hazırki vaxtda başa düşmə və qərar qəbuletmə üçün zəruri məlumatları özündə saxlaması, aktual olması başa düşülür. Belə ki, vaxt keçdikçə informasiya öz aktuallığını itirə bilər. Başqa sözlə, vaxtında daxil olmayan informasiya sonradan öz əhəmiyyətini itirə, lazımsız ola bilər. Əgər informasiya dəyişən şərtlər daxilində iş zamanı cari vəziyyətə uyğun olarsa, onda o, aktual hesab olunur. Vaxtında daxil olan informasiya zəruri fayda verə bilər.

Qiymətlik (əhəmiyyətlik, faydalı) – hər hansı məsələnin həll edilməsi üçün informasiyanın vaciblik dərəcəsi ilə müəyyən olunan xassəsidir. Başqa sözlə, informasiyanın qiymətliliyi onun köməyi ilə verilmiş məsələnin nə qədər effektiv həll edilməsi dərəcəsi ilə müəyyən olunur. Qiymətli informasiya eyni zamanda faydalı, kifayət qədər yeni, dolğun və doğru olmalıdır. Lakin bir çox hallarda informasiyanın dolğunluğu və doğruluğunun təmin edilməsi üzrə aparılan işlər onun köhnəlməsinə və təbii ki, faydasız olmasına gətirib çıxarır. Informasiyanın faydalılığı obyekt haqqında məlumatların qeyri-müəyyənliyinin azaldılması dərəcəsi ilə müəyyən edilir.

Aydınluq dedikdə informasiyanın təyinatı üzrə aidiyyəti şəxslər tərəfindən başa düşülən dildə ifadə olunması nəzərdə tutulur

1.4. İnformasiya sistemləri

Böyük həcmli informasiyanın toplanması, daxil edilməsi, emalı, saxlanması və ötürülməsi üçün informasiya texnologiyalarından, o cümlədən kompüter texnikasından istifadə olunur, bu və ya digər informasiya sistemi reallaşdırılır.

İnformasiya texnologiyaları – informasiyanın toplanması, emalı, saxlanması, ötürülməsi, təqdim edilməsi, əks olunmasını təmin edən, bir-biri ilə texnoloji zəncirdə birləşən üsullar, istehsal prosesləri və proqram-texniki vasitələr toplusudur. İnformasiya texnologiyalarının məqsədi informasiya ehtiyatlarının istifadəsi proseslərinin çətinliyinin azaldılmasından, onların etibarlığının və opertivliyinin yüksəldilməsindən ibarətdir. Ona görə də, tərfi aşağıdakı kimi vermək olar: informasiya texnologiyaları – informasiya ehtiyatlarının istifadəsi zamanı etibarlığı və opertivliyi yüksəltmək, yaranan çətinliyi azaltmaq məqsədilə zəruri informasiya proseslərinin yerinə yetirilməsini təmin edən və texnoloji zəncirdə birləşən üsullar və proqram-texniki vasitələr kompleksidir. Burada ***informasiya prosesləri*** dedikdə informasiyanın yaradılması, toplanması, işlənməsi, saxlanması, axtarışı və yayılması prosesləri nəzərdə tutulur.

İnformasiya ehtiyatları – informasiya sistemlərində (kitabxanalarda, arxivlərdə, fondlarda, verilənlər banklarında və s.) olan sənədlər və sənəd massivləri, habelə ayrıca mövcud olan sənədlər və onların massivləridir.

İnformasiya sistemi anlayışının geniş və dar mənada təriflərinə rast gəlinir. Geniş mənada ***informasiya sistemi*** (İS) dedikdə insanların zəruri informasiya ilə vaxtında təmin edilməsi üçün nəzərdə tutulmuş texniki, proqram vasitələrindən, təşkilati təminatdan və personaldan ibarət kompleks başa düşülür. “İnformasiya, informasiyalaşdırma və informasiyanın mühafizəsi haqqında” Azərbaycan Respublikasının Qanununda “***informasiya sistemi*** – informasiya texnologiyaları və sənədlərinin təşkilati və texniki qaydada, o cümlədən hesablama texnikasından istifadə edilməklə, nizamlanmış məcmusu” kimi şərh olunur.

Dar mənada İS dedikdə informasiyanın toplanması və emalının avtomatlaşdırılması üçün nəzərdə tutulmuş qarşılıqlı əlaqəli proqram-texniki vasitələr toplusu başa düşülür. İS-ə məlumatlar mənbələrdən daxil olur və gələcək istifadə üçün məlumat bazalarında saxlanılır. Zəruri hallarda müəyyən emal prosesini keçdikdən sonra məlumatlar istehlakçılara ötürülür.

Beləliklə, İS – informasiyanın toplanması, saxlanması, axtarılması, emal olunması və istifadəçilərə təqdim edilməsi üçün texniki (hesablama və kommunikasiya), proqram, linqvistik, təşkilati və metodoloji vasitələri özündə birləşdirən kompleksdir. İS qismində kitabxanaları, arxivləri, fondları, məlumat banklarını və bazalarını, internetin informasiya xidmətlərini və s. göstərmək olar.

Qeyd edilməlidir ki, informasiya sistemləri informasiya texnologiyaları ilə sıx əlaqəli olub, onlar üçün əsas mühit rolunu oynayır. Müvafiq informasiya texnologiyalarını bilmədən İS-in funksiyalarının reallaşdırılması mümkün deyildir. Lakin informasiya texnologiyaları İS sahəsindən kənar da mövcud ola bilər.

İlk informasiya sistemləri 1950-ci illərdə meydana gəlmişdir. Həmin illərdə İS hesabların emalı və əməkhaqqının hesablanması üçün nəzərdə tutulur, elektromexaniki mühasibat hesalama maşınlarında reallaşdırılırdı.

1960-cı illərdə İS-ə münasibət dəyişdi. İS-lərin köməyi ilə emal olunan informasiya bir çox parametrlər üzrə dövrü hesabatların aparılması üçün tətbiq olunmağa başladı. Bunun üçün çoxlu sayda funksiyaları yerinə yetirməyi bacaran geniş təyinatlı kompüter avadanlıqları tələb olunurdu.

1970-ci illərdə və 1980-ci illərin əvvəllərində İS qərar qəbul etmə prosesini dəstəkləyən və sürətləndirən idarəetmə nəzarəti vasitələri qismində geniş istifadə olunmağa başlandı.

1980-ci illərin sonuna doğru İS-lərin istifadəsi konsepsiyası yenidən dəyişdi. İS informasiyanın strateji mənbəyinə çevrilərək istənilən profilli təşkilatlarda bütün səviyyələrdə istifadə olunurdu. Həmin dövrlərdə İS təşkilatlara lazımi informasiyanı vaxtında təqdim etməklə xidməti fəaliyyətlərin, o cümlədən kommersiya fəaliyyətinin həyata keçirilməsinə, yeni mal və xidmətlərin yaradılmasına, satış bazarının və dəyərli tərəfdaşların axtarılmasına, məhsulların aşağı qiymətə istehsalına və s. yardım edirdi.

İS-də aşağıdakı informasiya prosesləri yerinə yetirilir:

- daxili və xarici mənbələrdən informasiyanın daxil edilməsi;
- sonradan istifadə olunması üçün informasiyanın saxlanılması;
- istifadəçinin sorğusuna görə informasiyanın axtarılması;
- giriş informasiyasının emalı və istifadəçi üçün rahat formada təqdim edilməsi;
- istifadəçiyə təqdim etmək və digər sistemə ötürmək üçün informasiyanın çıxarılması;
- yeni giriş informasiyasının korreksiyası üçün verilənlərin mütəxəssislər tərəfindən emalı (əks əlaqənin təşkili).

İS aşağıdakı əsas xassələrlə xarakterizə olunur:

- istənilən İS sistemlərin qurulmasının ümumi prinsipləri əsasında təhlil oluna, qurula və idarə edilə bilər;
- İS dinamik və inkişaf edən sistemdir;
- İS-in qurulması zamanı sistemli yanaşmadan istifadə olunur;
- İS kompüterdən, telekommunikasiya qurğularından ibarət olan, müasir texnologiyalar əsasında reallaşdırılan informasiya emalı sistemidir;

- İS-in çıxış məhsulu əsasında qərar qəbul edilən informasiyadır;
- İS-in işində insanın iştirakı sistemin mürəkkəbliyindən, verilənlərin növündən və toplusundan, həll edilən məsələlərin formallaşdırılma dərəcəsiindən asılıdır.

İstənilən İS-in strukturuna bir sıra təminat altsistemləri daxil ola bilər. Təminat altsistemlərinə aşağıdakılar aid etmək olar:

- informasiya təminatı altsistemi;
- texniki təminat altsistemi;
- riyazi təminat altsistemi;
- proqram təminatı altsistemi;
- təşkilati təminat altsistemi;
- hüquqi təminatı altsistemi.

İnformasiya təminatı – informasiyanın təsnifatlandırılması və kodlaşdırılmasının vahid sistemi, unifikasiya edilmiş sənədləşdirmə sistemləri, təşkilatda dövr edən informasiya axınları sxemləri, eləcə də verilənlər bazalarının qurulması metodologiyası məcmusudur.

Texniki təminat – İS-in işi üçün nəzərdə tutulmuş texniki vasitələr, eləcə də bu vasitələr əsasında reallaşdırılan müvafiq sənədlər və texnoloji proseslər kompleksidir.

Riyazi və proqram təminatı – İS-in məqsəd və vəzifələrinin reallaşdırılması, eləcə də texniki vasitələr kompleksinin normal fəaliyyəti üçün zəruri olan riyazi üsullar, modellər, alqoritmlər və proqramlar toplusudur.

Təşkilati təminat – İS-in işlənilib hazırlanması və istismarı prosesində əməkdaşların texniki vasitələrlə və öz aralarında qarşılıqlı fəaliyyətini nizamlayan üsullar və vasitələr toplusudur.

Hüquqi təminat – İS-in yaradılması, hüquqi statusu və fəaliyyətini müəyyən edən, informasiyanın alınması, çev-

rilməsi və istifadəsi qaydalarını nizamlayan hüquqi normalar toplusudur.

1.5. **İnformasiya sistemlərinin təsnifatı**

Emal olunan informasiyanın növünə görə informasiya sistemlərini iki sinfə bölmək olar:

- faktoqrafik informasiya sistemləri;
- sənədli informasiya sistemləri.

Faktoqrafik informasiya sistemləri (FİS) xüsusi formada təşkil edilən və formalaşdırılan verilənlər şəklində təqdim olunan faktoqrafik informasiya ilə işləyirlər. Yuxarıda qeyd olunduğu kimi, belə informasiya faktiki məlumat xarakteri daşıyır və verilənlər adlanırlar.

Faktoqrafik İS-in funksional bazasını verilənlər bazalarının idarə edilməsi üçün nəzərdə tutulmuş sistemlər təşkil edir. Ona görə də faktoqrafik informasiya sistemlərini, ənənəvi olaraq, **verilənlər bazalarının idarə edilməsi sistemləri** (VBİS) adlandırırlar.

VBİS-də verilənlər xüsusi strukturlaşdırılmış fayllarda – verilənlər bazalarında (VB) ayrı-ayrı yazılar – verilənlər toplusu şəklində saxlanılır. VB-də hər bir yazı konkret bir obyektə təsvir edir, həmin obyektin konkret xarakteristikası və ya parametri haqqında məlumatı özündə saxlayan bir və ya bir neçə hissədən, yəni sahədən ibarət olur.

Obyektlərin faktoqrafik təqdimatı onun xarakteristikalarının sadalanmasını, göstəricilərin müəyyən edilməsini, obyektin və ya prosesin vəziyyətinin faktiki təsvirini və s. nəzərdə tutur.

Ümumiyyətlə, faktoqrafik informasiya sistemləri məlumat-sorğu funksiyalarının reallaşdırılması və verilənlərin

emalı məsələlərinin həlli üçün istifadə olunur. *Məlumat-sorğu funksiyaları* dedikdə sorğu edilmiş informasiyanın dəqiqləşdirilməsi və ya konkretləşdirilməsi, habelə obyektlə bağlı daha çox məlumatın əldə olunması, rekvizitlərin, xarakteristikaların və göstəricilərin dəqiqləşdirilməsi və s. başa düşülür. *Verilənlərin emalı məsələləri* isə onların daxil edilməsi, dəyişdirilməsi, saxlanması, istifadə üçün aktualaşdırılması, seçilib götürülməsi ilə bağlı olan və kompüterlərin bazasında həyata keçirilən məsələləri özündə birləşdirir.

FİS-də toplanılan və emal edilən verilənlər, adətən, ciddi strukturlaşdırılmış olur, yəni kompüterdə saxlanması, emalı və istifadəsi onlar xüsusi şəkildə təşkil edilmiş struktura malik olurlar. Bu sistemlərdə verilənlər, əsasən, informasiya obyektlərinin parametrləri, göstəriciləri və digər xarakteristikaları haqqında məlumatların ədədi qiymətlərlə və ya sözlə konkret təsvirini özündə ehtiva edir.

Verilənlərin emal edilməsi və onlar üzərində tələb olunan əməliyyatların yerinə yetirilməsi üçün belə sistemlərdə bütün zəruri funksiyalar reallaşdırılır. Bu funksiyalara verilənlərin daxil edilməsi, dəyişdirilməsi, pozulması, axtarılması, seçilməsi, müqayisəsi, nizamlanması, çıxış (hesabat) formalarının tərtib edilməsi və s. daxildir.

Belə sistemlərə nümunə qismində kadrların qeydiyyatı, əməkhaqqının hesablanması, maddi qiymətlilərin və malların uçotu, bank hesablaşmaları və digər məsələlərin həlli üçün nəzərdə tutulmuş informasiya sistemlərini göstərmək olar.

Faktoqrafik sistemlərdə verilənlər dəqiq strukturlaşdırılmış və qabaqcadan müəyyən edilmiş formatda təqdim edildiyi üçün axtarışın aparılması məqsədilə sorğular dəqiq tərtib olunur, verilənlər və çıxış hesabatları baxış və istifadə üçün istifadəçilərə təsbit olunmuş formalarda təq-

dim edilir. Bu formalar sistemdə əvvəlcədən tərtib edildiyi üçün onların quruluşu məhdudlaşdırılmış olur.

Son illərdə təbii dildə olan mətn tipli sənədlərin, qrafik, audio-video və başqa formalarda olan verilənlərin işlənməsini, informasiyanın intellektual emalını yerinə yetirən avtomatlaşdırılmış informasiya sistemləri meydana gəlmişdir. Belə sistemlər strukturlaşdırılmış və formallaşdırılmış verilənlərlə deyil, strukturlaşdırılmamış, zəif strukturlaşdırılmış, hətta sistemləşdirilməmiş mətnləri, cədvəlləri, qrafikləri və audio-video verilənləri və s. özündə saxlayan mürəkkəb sənədlərlə işləyir. İS-in ikinci əsas sinfini təşkil edən bu sistemləri *sənədli informasiya sistemləri* (SİS) adlandırırlar.

Müasir sənədli informasiya sistemləri çox vaxt ayrı-ayrı müəlliflər tərəfindən müxtəlif formatlarda və kodlarda ayrı-ayrı məqsədlər üçün yaradılan və çox böyük həcmə malik olan informasiya ehtiyatlarını özündə birləşdirir. SİS-in verilənləri (ehtiyatları) və ya məlumatları qismində ASCII, Win, KOI, UNICODE və s. formatlarda, müxtəlif dillərdə hazırlanmış mətnlər, hipermətnlər, GIF, JPG, Photoshop, Corel, Paint formatlarında olan şəkillər, audio-video məlumatlar, Access, Oracle, Foxbase, Paradox sistemlərinin köməyi ilə yaradılmış verilənlər bazaları, Excel sistemində qurulmuş elektron cədvəllər, MS Word və ya Adobe Acrobat sistemlərində hazırlanmış mətnləri, qrafikləri, şəkilləri, cədvəlləri və s. özündə birləşdirən mürəkkəb sənədlər ola bilər.

SİS kitablar, dərsliklər, monoqrafiyalar, dissertasiyalar, avtoreferatlar, elmi məqalələr, qanunlar, kommersiya təklifləri, təsvirlər, təlimatlar və s. şəklində təbii dildə hazırlanmış sənədlərin tam mətnləri ilə işlədikləri üçün çox

vaxt onlara *tammətli informasiya sistemləri* (TİS) deyirlər.

SİS-də məlumatlar daha mürəkkəb struktura və xüsusiyyətə malik olduğuna görə bu sistemlərdə məlumatların təqdim olunması, saxlanması, emalı, seçilib götürülməsi, habelə axtarış sorğularının tərtib olunması və lazım olan informasiyanın axtarılması prinsipləri, üsul və mexanizmləri ənənəvi FİS-də reallaşdırılan üsul və mexanizmlərdən köklü surətdə fərqlənirlər.

Belə informasiya sistemləri ötən əsrin 70-ci illərindən inkişaf etməyə başlamışdır. O dövrlərdə yaradılmış INIS, INSPEC, STN, NTIS, MEDLAR kimi axtarış sistemləri geniş istifadə olunmuşdur. Rusiyada VİNİTİ tərəfindən təqdim olunan referativ verilənlər bazaları da öz praktiki əhəmiyyətinə görə bu gün də seçilir.

Strukturuna və paylanma dərəcəsinə görə də informasiya sistemlərini iki qrupa ayırmaq olar:

- lokal informasiya sistemləri;
- korporativ (paylanmış) informasiya sistemləri.

Tarixən ilk olaraq bir müəssisənin və ya təşkilatın faktoqrafik məlumatlarını özündə saxlayan *lokal informasiya sistemləri* meydana gəlmişdir. Onlar bir kompüterin bazasında qurulur, serverlə və ya digər qurğu ilə qarşılıqlı əlaqəsi olmayan informasiya sistemləridir. Lokal faktoqrafik informasiya sistemlərinə, bir qayda olaraq, əməkdaşlar, istehsal göstəriciləri, statistik məlumatlar, perspektiv planlar və s. haqqında verilənlər bazaları, yəni bütövlükdə müəssisənin fəaliyyəti haqqında məlumatlar daxil edilir. Lokal sənədli informasiya sistemlərinə isə müəssisə daxilində sənədlərin dövriyyəsi sistemlərini, yerinə yetirilən işlər haqqında hesabatları, əməkdaşların əsərlərini, el-

mi tədqiqatların nəticələrini və s. özündə saxlayan sistemləri aid etmək olar.

İnkişaf prosesində lokal informasiya sistemlərinin əksəriyyəti tədricən genişlənərək, onları yaradan, yəni onların sahibi olan bir müəssisənin çərçivəsindən kənara çıxmışdır. Bu onunla əlaqədar olmuşdur ki, belə sistemlərdə toplanan, həddən artıq böyük həcmə malik olan həmin müəssisənin fəaliyyəti ilə bağlı olan və ya bu fəaliyyət nəticəsində yaranan informasiya ehtiyatları eyni zamanda bu müəssisəyə aidiyyəti olmayan başqa şəxslərin və digər müəssisələrin maraq dairəsinə daxil olurdular. Beləliklə, informasiya sisteminin kənar şəxslər və təşkilatlar tərəfindən birgə istifadəsi zərurəti yaranırdı.

Məhz bu səbəbdən *korporativ (paylanmış) informasiya sistemləri* yaranmağa başlamışdır. Korporativ informasiya sistemləri bir müəssisə tərəfindən deyil, ümumi maraqları olan çoxlu sayda təşkilatların birgə iştirakı ilə yaradılır, yeniləşdirilir və istifadə olunur. Adətən, belə sistemlər elmi-texniki, iqtisadi, hüquq və s. sahələrdə informasiyanı özündə saxlayırlar. Aydınır ki, korporativ informasiya sistemlərinin inkişafına kompüter şəbəkələrinin, o cümlədən internetin yaranması əhəmiyyətli dərəcədə təsir göstərmişdir.

İS-i, həmçinin, aşağıdakı əlamətlərə görə də təsnif etmək olar.

İS-in qarşısında qoyulmuş vəzifələrə görə:

- idarəetmə hesabatlarını hazırlayan İS – əsasən verilənlərin emalına yönəlmiş olur;
- qərar qəbul edən İS – mümkün alternativ qərarlar işləyib hazırlayır. Qərarın qəbil edilməsi prosesi sistem tərəfindən təklif olunan alternativ qərarlardan birinin seçilməsi ilə nəticələnir. Belə İS-ə *ekspert sistemləri* deyirlər.

Avtomatlaşdırılma dərəcəsinə görə:

- əl ilə işləyən İS – informasiya emalının müasir texniki vasitələrinin tamamilə olmaması və bütün əməliyyatların insan tərəfində yerinə yetirilməsi ilə xarakterizə olunur;
- avtomatik İS – informasiyanın emalı üzrə bütün əməliyyatlar insanın iştirakı olmadan yerinə yetirilir;
- avtomatlaşdırılmış İS – informasiyanın emalı prosesində həm insanın, həm də texniki vasitələrin iştirakını nəzərdə tutur, lakin bu prosesdə əsas rolu kompüter oynayır. Texniki vasitələrin, o cümlədən kompüter texnikasının bazasında reallaşdırılan, informasiya proseslərini avtomatlaşdırmağa, informasiya ehtiyatlarını daşıyıcılarında saxlamağa və emal etməyə imkan verən informasiya sisteminə ənənəvi olaraq **avtomatlaşdırılmış informasiya sistemi** (AİS) deyirlər.

Hazırda bütün informasiya sistemlərində (faktoqrafik və ya sənədli, lokal və ya korporativ) informasiya ehtiyatlarının yaradılması, toplanması, saxlanması və emal olunması işləri müasir texniki vasitələrin, o cümlədən kompüter texnikasının köməyi ilə yerinə yetirilir və avtomatlaşdırılmış olur. Ona görə də müasir anlamda “informasiya sistemləri” dedikdə “avtomatlaşdırılmış informasiya sistemləri” nəzərdə tutulur.

Tətbiq sahəsinə görə:

- təşkilati idarəetmə İS – idarəetmə personalının fəaliyyətinin avtomatlaşdırılması üçün nəzərdə tutulur;
- texnoloji proseslərin idarə edilməsi İS – istehsal personalının fəaliyyətinin avtomatlaşdırılmasına xidmət edir;
- avtomatlaşdırılmış layihələndirmə İS – yeni texnika və texnologiyanın yaradılması zamanı mühəndis-la-

yihəçilərin, konstruktorların, arxitekturların, memarların və s. fəaliyyətinin avtomatlaşdırılması üçün nəzərdə tutulur. Onların əsas funksiyası mühəndis hesablarının, qrafik (çertyoj, sxem, plan) və layihə sənədlərinin, layihələndirilən obyektlərin modelləşdirilməsi və s. aiddir.

- inteqrasiya olunmuş (korporativ) İS – şirkətlərin əksər funksiyalarının avtomatlaşdırılması üçün istifadə olunur, layihələndirmədən tutmuş satışadək bütün fəaliyyət dövrünü əhatə edir.

İnformasiyanın istifadəsi xarakterinə görə:

- informasiya-təhlil sistemləri – müəyyən alqoritmə görə informasiyanın emalı (təhlili) əməliyyatlarını həyata keçirir. Bu növ sistemləri iki sinfə bölürlər: idarəedici və məsləhətverici İS;
- informasiya-axtarış sistemləri – informasiyanın daxil edilməsi, sistemləşdirilməsi, saxlanması, istifadəçinin sorğusuna görə informasiyanın tapılıb verilməsi funksiyalarını yerinə yetirir. Bura kitabxanalarda, dəmiryolu vağzallarında və hava limanlarında reallaşdırılan bilet satışı üzrə informasiya-axtarış sistemlərini aid etmək olar.

Ənənəvi faktoqrafik və sənədli AİS-lərdə, bir qayda olaraq, müxtəlif kriterilərə (sorğulara) görə axtarışı həyata keçirən mexanizmlər reallaşdırılmış olur. Belə sistemlərdə saxlanılan informasiyanın axtarılması, götürülməsi və emalı həmin sistemlərin daxili imkanları hesabına həyata keçirilir. Ona görə də bir çox hallarda belə sistemləri informasiya-axtarış sistemləri adlandırırlar.

Aydındır ki, faktoqrafik AİS-lər, əsasən, nəzərdə tutulmuş şəxslər və ya təşkilatlar tərəfindən istifadə olunduğuna görə sistemdə saxlanılan və emal olunan məlumatla-

rın strukturu, reallaşdırılan sorğuların formatı, dialoqun forması, axtarış mexanizmləri və s. onun istifadəçilərinə əvvəlcədən məlum olur. Belə sistemlər gündəlik fəaliyyətdə iş prosesində istifadə olunur.

Sənədli AİS-in istifadəçilərinin kateqoriyası və kontingenti əvvəlcədən dəqiq bəlli olmadığından onlarda bu sistemlərin iş prinsipləri, saxlanılan və emal edilən informasiyanın məzmunu, strukturu və xarakteri, eləcə də sorğuların formalaşdırılması və axtarışın aparılması mexanizmləri və s. barədə dəqiq məlumat olmaya bilər. Ona görə də bu növ informasiya sistemlərinin iş prinsipi, axtarış imkanları və s. haqqında məlumat onun istifadəçilərinə qabaqcadan çatdırılmalıdır. Əksər hallarda belə sistemlərdə istifadəçilərə kömək modulu reallaşdırılır.

Sənədli informasiya sistemləri internet şəbəkəsinin bazasında da müvəffəqiyyətlə yaradılır və fəaliyyət göstərirlər. Nümunə qisminə klassik sistemlər kimi tanınan Gopher, Wais, Archie, Whois, FTP, Listserv və digər informasiya-axtarış sistemlərini göstərmək olar.

Bundan əlavə, hal-hazırda mövcud olan bütün faktoqrafik və sənədli avtomatlaşdırılmış informasiya sistemləri, praktiki olaraq, tədricən internet şəbəkəsinə qoşulur və ümumi istifadə ehtiyatları kimi təşəkkül tapırlar. Burada şəxsi, kommersiya və digər məxfi xarakterli informasiya bazaları istisna təşkil edirlər. Əgər bu cür avtomatlaşdırılmış informasiya sistemləri internet şəbəkəsinə qoşulursa, onda onların qorunması üçün əlavə tədbirlərin görülməsi tələb olunur. Belə ki, bu sistemlərə giriş ciddi parol və digər təhlükəsiz daxilolma üsulu vasitəsilə təmin edilməlidir.

Web-texnologiyanın, o cümlədən HTML (HyperText Markup Language – hipermətnlərin təqdim edilməsi dili) dilinin və HTTP (HyperText Transfer Protocol) protoko-

lunun meydana gəlməsi ilə əlaqədar internetdə ilk olaraq informasiya axtarışı imkanları olmayan sənədli informasiya sistemləri yaranmağa başladı. Belə sistemlər, əsasən, böyük həcmdə informasiya ehtiyatlarını özündə saxlayır, onları istifadə etməyə, baxmağa, köçürməyə və çap etməyə, bir informasiya ehtiyatından digərinə keçməyə imkan verir.

Bu növ sistemləri sadəcə *internetin informasiya sistemləri* və ya *internetin informasiya ehtiyatları* adlandırırlar. Belə sistemlərdə, bir qayda olaraq, məna təhlilinə əsaslanan intellektual informasiya axtarışı funksiyaları reallaşdırılmır, lakin əksər sistemlərdə istifadəçilərə daxili ehtiyatlar üzrə axtarış aparmaq imkanı təklif olunur.

Ümumiyyətlə, *internetin informasiya ehtiyatları* dedikdə faktiki, sənədli və digər informasiya ehtiyatlarını özündə saxlayan, eləcə də internetə qoşulmaq imkanı olan və ya onun vasitəsilə öz ehtiyatlarına giriş təqdim edən istənilən kompüterləşdirilmiş (avtomatlaşdırılmış) informasiya sistemləri nəzərdə tutulur.

İnformasiya sistemlərinin yaradılmasının məqsədi isə müəyyən istifadəçi dairəsinin informasiya tələbatlarının ödənilməsi məqsədilə informasiyanın toplanması, saxlanması, axtarışı, emalı və ötürülməsi imkanlarının reallaşdırılmasından ibarətdir.

İnternetin informasiya ehtiyatları üzrə axtarış aparmaq üçün web-texnologiya üzərindən reallaşdırılan, xüsusi saytlar olan informasiya-axtarış sistemlərindən istifadə olunur.

II FƏSİL

İNTERNETİN İNFORMASIYA EHTİYATLARI VƏ XİDMƏTLƏRİ

İnternet – qlobal informasiya şəbəkəsi

İnternet ünvanlar

İnternetin informasiya ehtiyatları

(WWW, E-mail, FTP, Gopher, Telnet, BBS, Usenet, interaktiv
xidmətlər)

**İnternet şəbəkəsinin informasiya ehtiyatlarının
xüsusiyyətləri**

**İnformasiya ehtiyatlarının məzmunlarının təhlilinə
əsaslanan empirik paylanma qanunları**

(Zipf, Bredford, Pareto qanunları)

İnternetdə istifadə olunan əsas fayl formatları

2.1. İnternet – qlobal informasiya şəbəkəsi

İnternet – transmilli kompüter şəbəkələrinin birləşməsi olub, IP-protokolu və verilənlərin paketlərinin marşrutlaşdırılması bazasında qurulmuş beynəlxalq şəbəkədir. İnternet müxtəlif protokollarla işləyir, bütün növ mümkün kompüterlərin birləşməsini təmin edir, telefon naqilləri, optik kabellər, radiomodemlər və sputniklər vasitəsilə məlumatların ötürülməsini həyata keçirir.

Eyni zamanda qeyd edilməlidir ki, internet – müxtəlif növ kompüter şəbəkələrini özündə birləşdirən qlobal informasiya şəbəkəsidir. Başqa sözlə, internet – ən yeni texnologiyalara əsaslanan, geniş spektrdə informasiya və kommunikasiya ehtiyatlarına malik olan qlobal informasiya fəzasıdır. İnternet çox böyük həcmdə məlumatları özündə saxlayır, WWW və digər verilənlərin mübadiləsi sistemləri üçün fiziki baza rolunu oynayır. İnternet şəbəkəsini bəzən beynəlxalq şəbəkə, bəzən qlobal şəbəkə, bəzən isə sadəcə şəbəkə adlandırırlar.

Hazırda internet dedikdə fiziki kompüter şəbəkəsi yox, dünya hörümçək toru (WWW) üzərində formalaşmış informasiya fəzası nəzərdə tutulur.

İnternetin yaranma tarixi 2 sentyabr 1969-cu il hesab olunur. Həmin tarixdə ABŞ-ın Pespektiv müdafiə elmi-tədqiqat işləmələri üzrə Agentliyin (DARPA) təklifinə əsasən ABŞ Müdafiə Nazirliyinin maliyyə dəstəyi ilə həyata keçirilən layihə çərçivəsində ARPANET (Advanced Research Projects Agency Network) adlanan ilk kompüter şəbəkəsi istifadəyə verilmişdi. Bu layihə kritik şəraitlərdə informasiyanın mühafizəsini təmin etmək məqsədilə həyata keçirilmişdi. Həmin vaxt bu şəbəkə layihənin işti-

rakçıları olan ABŞ-ın 4 elmi təşkilatını birləşdirmişdi. ARPANET şəbəkəsinin ilk serveri (əməli yaddaşı 24 Kb olan Honeywell DP-516 kompüterü) həmin tarixdə Kaliforniya Universitetində quraşdırılmışdı. Bundan sonra ARPANET çox fəal şəkildə inkişaf etməyə və genişlənməyə başladı. Onun imkanlarından müxtəlif elm sahələrində fəaliyyət göstərən alimlər istifadə edirdilər.

Aparılan tədqiqatların nəticəsində 1971-ci ildə şəbəkə ilə elektron poçtun göndərilməsi üçün ilk proqram təminatı işlənilib hazırlandı və beləliklə, qısa zamanda çox məşhurlaşan elektron poçtu xidməti meydana gəldi. 1970-ci illər ərzində bu şəbəkə əsasən elektron poçtun göndərilməsi üçün istifadə olunurdu. Məhz həmin dövrdə poçt göndərilməsi siyahıları, xəbər qrupları və elanlar lövhələri yarandı.

1973-cü ildə transatlantik telefon kabeli vasitəsilə Böyük Britaniya və Norveçdən olan ilk xarici təşkilatlar bu şəbəkəyə qoşuldular. Beləliklə, o, beynəlxalq şəbəkə statusu qazandı.

1970-ci illərin sonuna doğru verilənlərin ötürülməsi protokolları inkişaf etməyə başladı. Bu protokollar 1982-83-cü illərdə standartlaşdırıldı. 1983-cü ildə ARPANET NCP protokolundan TCP/IP protokoluna keçdi. Həmin ildən ARPANET şəbəkəsi internet adlanmağa başladı. TCP/IP indiyə qədər internetdə kompüter şəbəkələrinin birləşməsi üçün uğurla tətbiq olunur. Bir il sonra domen adlar sistemi (Domain Name System – DNS) işlənilib hazırlandı.

1984-cü ildə ABŞ Milli elm fondu (National Science Foundation – NFS) tərəfindən NFSNet adlanan universitetlərarası geniş şəbəkə yaradıldı. NFSNet şəbəkəsi ARPANET şəbəkəsinə nisbətən daha böyük buraxıcılıq qabiliyyətinə və sürətə malik idi. Sonradan 1990-ci ildə ARPANET şəbəkəsi fəaliyyətini dayandırdı.

1988-ci ildə IRC (Internet Relay Chat) protokolu işlə-nib hazırlandı və internetdə real vaxtda ünsiyyətə imkan verən internet-söhbət xidməti reallaşdırıldı.

1990-cı ildə internetə telefon xətti vasitəsilə nömrə yığmaqla (dial-up access) ilk qoşulma həyata keçirildi.

İnternetin əsl çiçəklənməsi WWW (World Wide Web – dünya hörümçək toru) xidmətinin yaranması ilə başladı. Bu xidməti sadəlik naminə çox vaxt web adlandırırlar. WWW konsepsiyasının əsası CERN (fransızca: Conseil Europee pour Recherche Nucleaire – Nüvə Tədqiqatları üzrə Avropa Şurası) mərkəzində 1989-cu ildə britaniyalı alim Tim Berners-Li tərəfindən qoyulmuş, sonrakı iki il ərzində isə HTTP protokolu, HTML dili və URI identifikatoru işlə-nib hazırlanmışdır. Beləliklə, 1991-ci ildə WWW xidməti, 1993-cü ildə isə NCSA Mosaic adlanan məşhur web-brauser proqramı internetdə ümumi istifadə-yə verildi və bu xidmət məşurlaşmağa başladı.

Hazırda internet həm informasiya mənbəyi, həm də in-formasiya mübadiləsi vasitəsi rolunda çıxış edir. 1995-ci ildə WWW internetin əsas informasiya təchizatçısına çevrildi.

Bu gün internet istifadəçilərinin sayı astronomik sürət-lə artır. Açıq informasiya mənbələrinin statistik məlumat-larına görə, hazırda internet dünyanın 200-ə yaxın ölkə-sini əhatə edir. İnternet şəbəkəsinə 1998-ci ildə təxminən 143 milyon, 2010-cu ildə 1 milyard 600 milyon istifadəçi qoşulmuşdusa, 2013-cü ilin sonunda bu göstərici 2,8 mil-yardı (dünya əhalisinin 39 faizini) ötüb keçmişdir.

2013-cü ilin dekabrına olan məlumatlara görə dünyada qitələr üzrə internet istifadəçilərinin sayı aşağıdakı kimi olmuşdur:

- Asiyada – 1,27 milyard (əhalinin 31,7%-i);
- Avropada – 566 milyon (əhalinin 68,6%-i);

- Latın Amerikasında – 302 milyon (əhalinin 49,3%-i);
- Şimali Amerikada – 300 milyon (əhalinin 84,9%-i);
- Afrikada – 240 milyon (əhalinin 21,3%-i);
- Avstraliyada – 24,8 milyon (əhalinin 67,5%-i).

Dünya ölkələri arasında internet istifadəçilərinin sayına görə müəyyən edilmiş siyahıda ilk on yeri aşağıdakı ölkələr tutur:

- Çin – 620 milyon (əhalinin 45,8%-i);
- ABŞ – 268,5 milyon (əhalinin 84,2%-i);
- Hindistan – 195 milyon (əhalinin 15,8%-i);
- Braziliya – 109,8 milyon (əhalinin 54,2%-i);
- Yaponiya – 109,6 milyon (əhalinin 86,2%-i);
- Rusiya – 87,5 milyon (əhalinin 61,4%-i);
- Almaniya – 69,8 milyon (əhalinin 86,2%-i);
- Nigeriya – 67,3 milyon (əhalinin 38%-i);
- Böyük Britaniya – 57,3 milyon (əhalinin 89,8%-i);
- Fransa – 55,2 milyon (əhalinin 83,3%-i).

Statistik məlumatlara görə Azərbaycan Respublikasında internet şəbəkəsinin xidmətlərindən istifadə edənlərin sayı artıq 5,7 milyonu (əhalinin 58,7%-i) ötüb keçmişdir.

İnternetdə mövcud olan müxtəlif növ şəbəkələrin qarşılıqlı fəaliyyətinin, şəbəkə qovşaqları arasında əlaqənin və informasiya mübadiləsinin təmin edilməsi məqsədilə bu gün onlarla protokollarla reallaşdırılmışdır. Bütövlükdə internetin protokollar sistemi “TCP/IP protokolları” adlanır.

İnternetin protokolları açıq sistemlərin qarşılıqlı əlaqə (OSI) modelinin müxtəlif səviyyələrində işləyir, onları daha çox istifadə olunan protokolları OSI modelinin səviyyələri üzrə aşağıdakı kimi təsnif etmək olar: tətbiqi səviyyəsi – FTP, HTTP, HTTPS, SMTP, POP3, Telnet, DNS, IMAP, LDAP, BGP, SNMP, SSH; seans və təqdimat səviyyələrində – SSL, TLS; nəqliyyat səviyyəsində – TCP,

UDP; şəbəkə səviyyəsində – EIGRP, ICMP, IGMP, IP, IS-IS, OSPF, RIP; kanal səviyyəsində – Arcnet, ATM, Ethernet, Frame relay, HDLC, PPP, L2TP, SLIP, Token ring.

Verilənlər şəbəkədə hissələrə bölünərək ötürülür. İnternet vasitəsilə ötürülə bilməsi üçün həmin hissəyə alan kompüterin ünvanı, paketin sıra nömrəsi və digər məlumatlardan ibarət olan idarəedici informasiya (başlıq) əlavə edilir. İdarəedici informasiya (başlıq) əlavə edilmiş verilənlər *paket* adlanır.

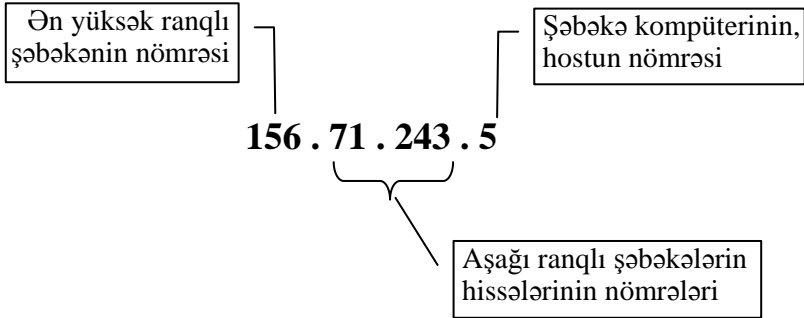
Paketlərin formatı və həcmi, eləcə də onların internetdə ötürülməsi və yığılması ardıcılığı TCP/IP (Transmission Control Protocol/Internet Protocol – ötürmənin idarə edilməsi protokolu / şəbəkələrarası protokol) protokolları ilə müəyyən edilir. TCP protokolu şəbəkədə iki kompüter arasında rabitə seansının təşkilini, IP isə marşrutlaşmaya, yəni paketlərin lazımı ünvana çatdırılmasını təmin edirlər.

Hazırda internet şəbəkəsi IP (Internet Protocol) protokolunun və verilənlər paketlərinin marşrutlaşdırılması texnologiyasının köməyi ilə müxtəlif arxitekturalara və topologiyaya malik olan minlərlə korporativ, elmi, hökumət, özəl və fərdi kompüter şəbəkələrini özündə birləşdirir.

2.2. İnternet ünvanlar

Verilənlərin ötürülməsi zamanı problemlərin yaranmaması üçün internetdə qoşulmuş hər bir kompüter unikal IP-ünvana malik olmalıdır. IP-ünvan bir-birindən nöqtə ilə ayrılan dörd ədəd ardıcılığından ibarət olur. Bu ədədlərin hər biri bir bayt uzunluğunda olur və 0-dan 255-dək qiymət ala bilər. Ən soldakı ədəd ən yüksək rəngli şəbəkənin nömrəsi olur. Sonrakı iki ədəd nisbətən aşağı rəngli

şəbəkələrin hissələrini, sonuncu ədəd isə konkret şəbəkə kompüterini (hostu) müəyyən edir (şək.2.1).



Şək.2.1. IP-ünvan

Ümumi halda, *host* dedikdə digər avadanlıqlarla ünsiyyət üçün TCP/IP protokollarını istifadə edən istənilən qurğu başa düşülür.

IP-ünvanlar çətin yadda saxlanıldığı üçün onların əvəzinə şərti simvolik adlardan istifadə olunması təklif olunmuşdur. Belə adlara domen adlar deyirlər. Domen adlar da bir-birindən nöqtələrlə ayrılan hissələrdən ibarət olur. Bu hissələrin sayı müxtəlif ola bilər. IP-ünvanlarda yüksək rəngli hissə solda, domen adlarda isə sağ tərəfdə yerləşir. Domen adın ən sağında yerləşən hissəsi birinci səviyyəli domenin adı və ya identifikatoru adlanır. Bu hissə, bir qayda olaraq, ölkələrin domen adları (*ru* – Rusiya, *pl* – Polşa, *tr* – Türkiyə, *az* – Azərbaycan və s.) və ya istifadə üçün ABŞ-a məxsus olan və beynəlxalq səviyyədə istifadə edilən domen adlar (*com* – kommersiya təşkilatları, *edu* – təhsil müəssisələri, *gov* – dövlət təşkilatları, *mil* – hərbi təşkilatlar, *net* – fəaliyyəti şəbəkə ilə

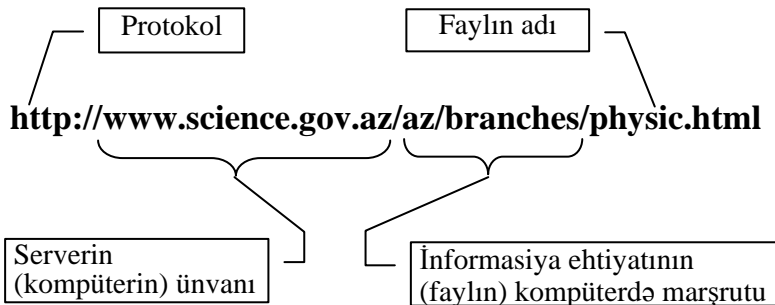
bağlı olan təşkilatlar, **org** – qeyri-kommersiya təşkilatları, **int** – beynəlxalq təşkilatlar və s.) olur.

İnternetdə işləyən zaman istifadəçilərin daxil etdiyi simvolik ünvan (domen ad) DNS (Domain Name System – Domen adlar sistemi) adlanan xüsusi xidmət tərəfindən IP-ünvana çevrilir.

2014-cü ilin əvvəlinə olan məlumata görə yuxarı səviyyəli domenlərin sayı 400-ü, yuxarı domen səviyyəsində qeydiyyatdan keçmiş domen adların sayı 280 milyonu, yuxarı domen səviyyəsi ölkə kodları olan qeydiyyatdan keçmiş domen adların sayı 150 milyonu ötüb keçmişdir.

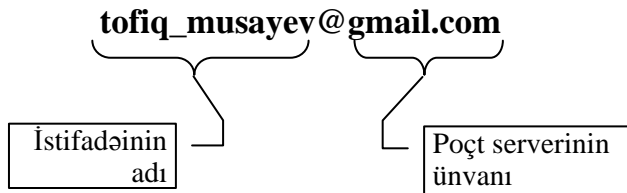
Qeyd olunduğu kimi, internetə qoşulmuş hər bir kompüter IP-ünvan və ya simvolik ünvan şəklində təqdim olunan unikal ünvanı malik olur. Lakin internetə qoşulmuş kompüterdə olan konkret informasiya ehtiyatını (məlumatı, sənədi və s. saxlayan faylı) əldə etmək üçün onun yerləşdiyi tam marşrutu və adını göstərmək lazımdır.

Bununla yanaşı, internet vasitəsilə həmin informasiya ehtiyatına müraciət etmək üçün əlavə olaraq ona daxil olmaq üçün zəruri olan protokol da göstərilməlidir. İnformasiya ehtiyatının adını, kompüterdəki marşrutunu, həmin komüterin internet ünvanını və müvafiq protokolun adını özündə birləşdirən simvollar ardıcılığı həmin **ehtiyatın internetdə bilavasitə göstəricisi** kimi müəyyən olunur. Bu göstərici informasiya ehtiyatının İnternedə unikal ünvanı olub URL (Universal Resource Locator – ehtiyatların universal göstəricisi) adlanır. Ona bəzən URL-ünvan, bəzən ehtiyatın internet ünvanı, bəzən isə sadəcə ünvan deyirlər. Tam URL-ünvan aşağıdakı şəkildə olur: protokol://serverin-ünvanı/marşrut/faylın adı (şəkl.2.2).



Sək.2.2. URL-ünvan

Qeyd olunmalıdır ki, elektron poçtu ünvanı URL-ünvanlardan fərqlənir. Belə ki, elektron poçt ünvanı “@” işarəsi ilə ayrılan iki hissədən ibarət olur. Bu simvoldan sol tərəfdə istifadəçinin (elektron poçtunun abonentinin) adı, sağ tərəfdə isə onun “poçt qutusunun” yerləşdiyi poçt serverinin domen adı (ünvanı) göstərilir: username@serverin adı (şək.2.3).



Sək.2.3. Elektron poçtu ünvanı

İstifadəçinin adı serverdə qeydiyyat zamanı onun özü tərəfindən seçilən və daxil edilən şərti addır. Məsələn, tofiq_musayev@gmail.com.

2.3. İnternetin informasiya ehtiyatları

Yeni informasiya texnologiyalarının kütləvi tətbiqi illərlə toplanmış ənənəvi informasiya ehtiyatlarının geniş miqyasda elektron formaya keçirilməsinə və tamamilə yeni şəkildə olan (o cümlədən multimediyalı) informasiya ehtiyatlarının yaranmasına gətirib çıxarmışdır. İnformasiya ehtiyatlarının elektron formada təqdim olunması onların daha geniş yayılmasına, əlverişli və səmərəli istifadəsinə imkan yaradır.

Hazırda beynəlxalq aləmdə informasiya fəzasında həm ixtisaslaşdırılmış, həm də ümumi təyinatlı informasiya sistemlərinin yaradılmasında internet texnologiyasının istifadəsi üstünlük təşkil edir.

Məlumdur ki, strukturuna görə internet nəhəng, lakin kifayət qədər çevik informasiya şəbəkəsidir. O, müxtəlif növ informasiya ehtiyatlarının yerləşdirilməsi üçün geniş imkanlar açır və bu ehtiyatlara olduqca sadə giriş interfeysi təqdim edir. Bu gün internet şəbəkəsinin bazasında onun istifadəçilərinə müxtəlif informasiya xidmətləri təklif edən çoxsaylı informasiya sistemləri müvəffəqiyyətlə fəaliyyət göstərir.

Müasir cəmiyyətin başlıca informasiya mənbələrindən və təchizatçılarından biri olan internet şəbəkəsinin əsas informasiya və telekommunikasiya xidmətləri aşağıdakılardır:

- WWW (World Wide Web) – dünya hörümçək toru;
- E-mail (Electronic mail) – elektron poçtu, məlumatların göndərilməsi siyahıları;
- FTP (File Transfer Protocol) – faylların ötürülməsi protokolu;
- Gopher – iyerarxik menyu;

- Telnet –uzaq məsafədən idarəetmə və sistemə daxilolma;
- BBS (Bulletin Board System) – elektron elan lövhələri
- Usenet – telekonfrans (xəbərlər) sistemi;
- İnternetin interaktiv informasiya xidmətləri;
- İnternet-telefon, internet-radio və internet-televiziya.

Növbəti bölmələrdə internetin yuxarıda adları çəkilən xidmətləri haqqında məlumat verilir. Qeyd olunmalıdır ki, burada həmin xidmətlər proqram-texniki və texnoloji baxımdan deyil, internet mühitində informasiya təminatını həyata keçirən xidmət kimi baxılırlar.

2.3.1. WWW – dünya hörümçək toru

WWW – internetə qoşulmuş müxtəlif kompüterlərdə yerləşən və bir-biri ilə hiperəlaqələrə malik olan sənədlərə girişi təqdim edən paylanmış informasiya sistemidir. WWW abreviaturası ilə yanaşı, W3 yazılışına da rast gəlinir.

İlk hipermətn layihəsi 1989-cu ildə CERN-də proqram təminatı üzrə məsləhətçi olan Tim Berners-Li tərəfindən təklif olunmuşdur. O, həmçinin HTTP, URI/URL və HTML texnologiyalarının da müəllifidir.

WWW (onu çox vaxt web-texnologiya adlandırırlar) hipermətn texnologiyası üzərində qurulmuş geniş və rəngarəng imkanlara malik informasiya xidmətidir. İnternetin digər xidmətləri arasında web-texnologiya başlıca yer tutur. Belə ki, internet istifadəçilərinin təxminən 79%-i bu xidmətdən istifadə edir. Bu xidmətin imkanları, eləcə də onun informasiya ehtiyatlarının həcmi astronomik sürətlə artır.

Hazırda dünya hörümçək toru yüz milyonlarla web-serverlərdən, milyardlarla web-səhifədən ibarətdir. 2014-cü ilin sentyabrına olan məlumata görə dünyada web-serverlərin sayı 1 milyardı ötüb keçmişdir.

Web-texnologiya istifadə üçün çox rahat və əlverişlidir. O, adi mətnlərin, hipermətnlərin, cədvəllərin, qrafik və audio-video multimediyalı məlumatların, proqram kodlarının və digər informasiya ehtiyatlarının internet şəbəkəsində yerləşdirilməsinə, saxlanılmasına, istifadəsinə, alınmasına və ötürülməsinə imkan verir.

Web-texnologiyanın reallaşdırılmasında dörd əsas komponentdən istifadə olunur:

- HTML (Hyper Text Markup Language) – hipermətnlərin təqdim edilməsi dili;
- URL (Universal Resource Locator) – ehtiyatların (ehtiyatların) internetdə ünvanlarının müəyyənləşdirilməsinin universal üsulu;
- HTTP (Hyper Text Transfer Protocol) – hipermətnli məlumatların mübadiləsi protokolu;
- CGI (Common Gateway Interface) – şlüzlərin universal interfeysi.

Web-serverlərdə informasiya ehtiyatları öz aralarında hipermətn istinadları ilə əlaqəli olan sənədlərdən ibarət olurlar. Hipermətn texnologiyası əsasında hazırlanan belə sənədlər *web-səhifə* adlanır. İnternet şəbəkəsinin bir qovşağında yerləşən və hiperistinadlarla əlaqəli olan web-səhifələr toplusu *web-server* və ya *web-sayt* adlanır. *Hipermətn* – özündə digər mətnlərə, qrafik, audio-video və s. informasiyaya əlaqələri (hiperistinadları) saxlayan mətnidir.

Web-səhifələrin yaradılması üçün istifadə olunan HTML dili web-səhifələrə istənilən növ məlumatı daxil etməyə, səhifənin bir hissəsindən digər hissəsinə, eləcə də

bir web-səhifədən digər web-səhifəyə istinadlar (hiperistinad və ya hiperkeçidlər) qoymağa və digər funksiyaları yerinə yetirməyə imkanı verir.

Hiperistinadlar web-səhifələrdə hər hansı sözə, sözlərə, qrafik elementlərə, şəkillərə və digər növ elementlərə bağlanır və həmin element haqqında daha geniş məlumatı əks etdirən hissəyə və ya digər web-səhifəyə keçidi təmin edir. Web-səhifələr arasında bu cür əlaqələr yalnız eyni bir saytda olan sənədləri deyil, həmçinin, müxtəlif saytlarda, hətta dünyanın müxtəlif ölkələrində yerləşən saytlarda olan məlumatları bir-biri ilə əlaqələndirə bilər. Bu cür yaradılmış "əlaqə telləri" son nəticədə hörümçək toruna oxşar web-əlaqələr torunu əmələ gətirir. Məhz bu baxımdan web-fəzanı dünya hörümçək toru adlandırırlar. Belə hiperəlaqələr üzrə bir informasiya ehtiyatından digərinə keçid web-fəzada naviqasiyanı, yəni informasiya ehtiyatları üzrə "gəzintini" təmin edir.

Web-saytlara qoşulmaq, saytlarda olan web-səhifələrə baxmaq və onlar arasında naviqasiyanı həyata keçirmək üçün *brauzer* (browser) adlanan xüsusi proqramlardan istifadə olunur. Brauzer proqramları bir sıra funksiyaları yerinə yetirir. Əvvəla, brauzer web-serverin müştəri proqramı funksiyasını yerinə yetirir. O, web-serverlə əlaqə yaradır və ondan web-səhifəni tələb edir. İkincisi, brauzer kompüterdə web-fəzanın pəncərəsi rolunda çıxış edir, internetdən alınmış sənədləri müəyyən olunmuş qaydada ekranda əks etdirir. Brauzerlərin digər funksiyaları qismində açılmış səhifələrin siyahısının saxlanması, axtarışın sadələşdirilməsi, faylların yüklənməsi və s. göstərmək olar. Bu gün Microsoft Internet Explorer, Netscape Navigator, Opera, Google Chrome, Mozilla və s. kimi brauzerlərdən daha çox istifadə olunur.

İnternetdə web-portal anlayışına da tez-tez rast gəlinir. Onlar təyinat və funksional baxımdan adi saytlardan bir qədər fərqlənirlər.

Web-portallar – təyinatlarına uyğun olaraq web-saytlar çoxluğuna, bu saytlarda yerləşən informasiya ehtiyatlarına, eləcə də müxtəlif internet-xidmətləri realizə edən proqramlara birbaşa girişi təmin edən çoxfunksiyalı web-saytlardır.

Web-portallarda əksini tapmış serverlər, saytlar, informasiya ehtiyatları və proqramlar hər hansı bir sistemə və ya müxtəlif sistemlərə aid ola bilər. Onlar özündə saxladıkları sənədlərin və verilənlərin mövzularına və ya digər əlamətlərinə görə xüsusi şəkildə seçilirlər. Adətən, istifadəçiləri (müşəriləri) özlərinə cəlb etmək üçün web-portallarda müxtəlif maraqlı funksiyalar reallaşdırılır. Portalarda, əsasən, axtarış, mövzu kataloqları, maliyyə indeksləri, hava haqqında məlumatlar və digər sorğu xidmətləri üstünlük təşkil edir.

İnformasiya xidmətinin ixtisaslaşmasına görə portalları iki kateqoriyaya bölürlər: üfüqi portallar və şaquli portallar.

Üfüqi portal – müxtəlif xidmətlər çoxluğunu təqdim edən, geniş mövzular çoxluğunu özündə birləşdirən, maksimal dərəcədə geniş istifadəçi auditoriyası üçün nəzərdə tutulan və onların mümkün maraqlarını əhatə edən portaldır. Belə portallar müxtəlif funksiyaları özündə birləşdirir, müxtəlif məzmunlu informasiya ehtiyatlarını və xidmətlərini təklif edirlər.

Şaquli portal – müəyyən mövzu sahəsində ümumi maraqları olan istifadəçilər üçün müxtəlif xidmətlər təqdim edən, müvafiq fəaliyyət sahəsinin tam əhatə olunmasına yönəlmiş dar mövzu yönümlü portaldır.

İnternetdə web-saytların işinin təşkili, onlara qoşulma və saytların məzmununa baxma HTTP protokolu vasitəsilə həyata keçirilir. Ona görə də web-saytların ünvanlarını yazarkən onların əvvəlinə “http://” əlavə olunur, lakin bu mütləq şəkildə tələb olunmur.

Hazırda insanlar internetin xidmətlərinə yalnız kompüter vasitəsilə deyil, eləcə də digər elektron qurğular (mobil telefonlar, elektron qeyd kitabları, rəqəmsal televizorlar və s.) vasitəsilə giriş əldə edirlər. Ona görə də kompüterlə yanaşı, mobil telefonlar, elektron qeyd kitabları, rəqəmsal televizorlar və s. üçün də brauzer proqramları işlənib hazırlanır və istifadə olunur.

Daha perspektivli və çox istifadə olunan elektron qurğu mobil telefondur. Son nəsil mobil telefonlar web-brauzer və elektron poçtu imkanları ilə təmin edilirlər. Mobil telefon və internet arasında məlumat mübadiləsi WAP (Wireless Application Protocol) protokolu vasitəsilə həyata keçirilir. Bu protokolun adına uyğun olaraq mobil qurğular üçün nəzərdə tutulan proqram əlavələri də WAP proqramlar adlanır.

İnternetin yuxarıda adları çəkilən informasiya xidmətləri web üzərindən reallaşdırılır. Belə ki, web-texnologiya internetin xidməti olmaqla yanaşı müxtəlif web-portalların, web-forumların, wiki layihələrin, internet mağazaların, auksionların, Facebook, Twitter, Vkontakte, Odnoklasniki və s. sosial şəbəkələrin, reklam və baner sistemlərinin, aktiv informasiya ehtiyatlarının, informasiya-axtarış sistemlərinin reallaşdırılması üçün baza rolunu oynayır.

İnternetin informasiya fəzasının yaranması və inkişafının ilk dövrü web-saytların dizayn stili olan **Web 1.0** üzərində qurulmuşdu. Həmin dövrdə internetin saytları, web əsaslı informasiya sistemləri yaranır və inkişaf edirdi. Bu

saytların informasiya kontenti (məzmunu) onların sahibləri və yaradıcıları tərəfindən həyata keçirilirdi. Saytların informasiya məzmunu onların sahiblərinin arzu, istək və məqsədlərindən, eləcə də qarşıya qoyduqları vəzifələrdən asılı olaraq formalaşırdı. İnternetin, o cümlədən həmin saytların istifadəçiləri bu prosesdə iştirak etmirdilər.

2005-ci ildə Web 2.0 platforması meydana gəldi. **Web 2.0** – şəbəkədə qarşılıqlı fəaliyyətlərin reallaşdırılması yolu ilə daha çox adam tərəfindən istifadə olunduqca keyfiyyətləri daha da yaxşılaşan sistemlərin layihələndirilməsi metodikasıdır. Başqa sözlə, Web 2.0 – web-ehtiyatların təşkili, reallaşdırılması və dəstəklənməsinə kompleks yanaşma olub, əsasında saytların məzmununun (kontentinin), o cümlədən informasiya materiallarının doldurulması və dəfələrlə yoxlanılması prosesinə istifadəçilərin cəlb edilməsi prinsipi durur. Mahiyyət etibarını ilə Web 2.0 istifadəçilərin özləri tərəfindən fəal surətdə inkişaf etdirilən və təkmilləşdirilən layihə və xidmətləri özündə birləşdirir. Onlara bloqlar, wiki-layihələr, sosial şəbəkələr, kitabxana 2.0 və s. layihələr aiddir.

Web 2.0 yanaşmasının istifadəsi əhəmiyyətli sayda yüksək istehlak keyfiyyətlərinə malik güclü internet-xidmətləri reallaşdırmağa imkan verdi. Bu isə eyni formalı (yeknəsəq) informasiya ehtiyatlarının yaranmasına və sonradan onların əksəriyyətinin qiymətliyiinin itməsinə gətirib çıxardı. Web 2.0 əsasında yaranan informasiya ehtiyatlarının məzmunu heç də həmişə tələb olunan səviyyədə olmur, bir çox hallarda peşəkar tələblərə cavab vermir, subyektiv xarakter daşıyır, həqiqəti əks etdirmir, ziddiyyətli məlumatları özündə saxlayır və s. Ona görə də Web 2.0 vasitəsilə saytların kontentinin doldurulması zamanı keyfiyyətli məqalələrin qeyri-peşəkarlar istifadəçilər

tərəfindən redaktə edilməsinə qadağa qoyulması, yeni daxil edilən məqalələrin peşəkar redaktorlar tərəfindən rəy verildikdən sonra saytda nəşr edilməsi ideyası formalaşmağa başladı. Bu da yeni platformaya keçid zərurətini doğurdu. Beləliklə, 2007-ci ildə Web 3.0 termini barədə internetdə məlumatlar yayılmağa başladı.

Web 3.0 – Web 2.0 texnoloji platforması bazasına peşəkarlar tərəfindən yüksəkkeyfiyyətli məzmun və xidmətlər yaratmağa imkan verən platformadır. Başqa sözlə, Web 2.0 mahiyyət etibarını ilə texnoloji platformadır, Web 3.0 isə bu texnoloji baza üzərindən peşəkarların köməyi ilə (istifadəçilərin daxil etdikləri məlumatları peşəkarların süzgəcindən keçirməklə) yüksək keyfiyyətli, maraqlı və faydalı məzmun (kontent) yaratmağa imkan verən platformadır.

Bloq – əsas məzmunu müntəzəm olaraq istifadəçilər tərəfindən daxil edilən, mətn, şəkil və mültimediya məlumatlarını özündə birləşdirən yazılardan ibarət web-saytdır. Bloq sözü ingiliscə “*Web log*” (*Blog*) sözlərindən əmələ gəlmişdir, internet gündəlik və ya hadisələrin internet jurnalı mənasını daşıyır. Bloqları yaradan və aparən şəxslər bloqçu (bloqqer) adlanır.

Bloqlara, bir qayda olaraq, müvəqqəti əhəmiyyətə malik olan qısa yazılar daxil edilir. Bu yazılar xronoloji baxımdan (daxil edilmə tarixinə görə) əks ardıcılıqla nizamlanır. Ən son daxil edilmiş yazı birinci olur. Ənənəvi gündəlikdən fərqli olaraq, bloqlarda olan yazılar ümumi (ictimai) istifadə təyinatlı olur və burada onların kənar istifadəçilər (oxucular) tərəfindən oxunması imkanı nəzərdə tutur. Kənar oxucular yazının müəllifi (bloqun sahibi – bloqqer) ilə müzakirəyə (polmikaya) qoşula bilər. Ona görə də bloqlarda müəyyən obyektlər, hadisələr, proseslər və s. haqqında məlumatlar, bilgilər toplanır.

Wiki layihə – məzmunu və strukturunu istifadəçilər tərəfindən müstəqil şəkildə dəyişdirilə bilən web-saytdır. Məzmunun və strukturun dəyişdirilməsi üçün alətlər həmin sayt tərəfindən təqdim olunur. Wiki saytlar istifadəçilərə adi web-brauzerdən istifadə etməklə istənilən səhifəni redaktə etməyə və ya yeni səhifə yaratmağa imkan verir. Bu məqsədlə mətnlərin strukturlaşdırılması, hiperistinadların qoyulması, formatlaşdırılması və s. imkanlar yarıdan wiki nişanlama “dilindən” istifadə olunur.

Wiki layihələr Web üzərindən olduqca böyük informasiya massivlərinin formalaşmasına gətirib çıxarmışdır (məsələn, Wikipedia, WikiLeaks). Lakin wiki saytlarda toplanan məlumatların doğruluğuna və həqiqiliyinə zəmanət vermək mümkün deyil. Ona görə də rəsmi mənbə kimi onlara istinad etmək tövsiyə olunmur.

Sosial şəbəkə – istifadəçilər arasında sosial qarşılıqlı əlaqələrin qurulması, əks etdirilməsi və təşkili üçün nəzərdə tutulmuş platforma, interaktiv (online) xidmət və ya web-saytdır. Sosial şəbəkələr aşağıdakı imkanların reallaşdırılması ilə xarakterizə olunur:

- istifadəçi tərəfindən özünün açıq və ya yarımaçıq profilinin yaradılması və idarə edilməsi;
- istifadəçi müəyyən əlaqəsi olan digər istifadəçilərin siyahısını şəbəkəyə verə və ya dəstəkləyə bilər;
- müxtəlif üzvlük kriterilərinə uyğun olaraq istifadəçi qruplarının formalaşdırılması;
- sistem daxilində istifadəçilər arasında mövcud olan əlaqələrin baxılması və yoxlanması.

Bu gün internetdə çoxlu sayda sosial şəbəkələr fəaliyyət göstərir. Daha məşhur sosial şəbəkələrin istifadəçilərinin sayına görə statistik göstəricilər 2014-cü ildə belə olmuşdur:

- Facebook – 1,4 milyard;
- YouTube – 1 milyard;
- WhatsApp – 600 milyon;
- Google+ – 540 milyon;
- Twitter – 500 milyon;
- LinkedIn – 260 milyon;
- V Kontakte – 230 milyon;
- Instagram – 200 milyon;
- Odnoklasniki – 205 milyon və s.

Web forum – web-saytın istifadəçilərinin qarşılıqlı ünsiyyətinin təşkili məqsədilə reallaşdırılan web-proqramdır. Web-forum – saytın qonağı (istifadəçisi - sayta daxil olan) tərəfindən öz məlumatını daxil etməyə və saxlamağı imkan verən bölməsidir. Bu zaman saxlanılan məlumatlar digər istifadəçilər tərəfindən də oxuna bilər.

Forumlar müzakirə üçün bölmələr toplusunu təqdim edir. İstifadəçilər bu bölmələrdə mövzular yaradırlar və həmin mövzular çərçivəsində müzakirələr aparırlar. Hər bir mövzu tematik qonaq kitabını xatırladır. Yəni mövzuya daxil olan bütün istifadəçilər (qonaqlar) orada öz məlumatlarını yazı bilərlər.

Ümumiyyətlə, web-forumun iyerarxiyası “bölmə - mövzu – məlumat” şəklində olur. Burada hər bir məlumat “müəllif-mövzu-məzmun-tarix/vaxt” şəklində informasiya daşıyır. Məlumat və ona cavablar mövzusunun qolunu (budağını) təşkil edir. Forumun ilkin müzakirə mövzusunda yayınma forumun qaydaları ilə qadağan olunur. Qaydalara riayət olunmasına müəyyən bölmə və ya mövzu çərçivəsində yad məlumatları redaktə etmək, başqa yerə daşımaq və pozmaq imkanlarına malik olan *moderatorlar* və *administratorlar* nəzarət edirlər.

Forumlarda məlumatlara girişin məhdudlaşdırılması əlahiddə dərəcədə çevik şəkildə həyata keçirilə bilər. Belə ki, bəzi forumlarda istənilən qonağa (iştirakçıya) yeni məlumatların oxunması və yaradılması imkanları verilir, digər forumlarda isə qabaqcadan qeydiyyat tələb olunur. Bəzən qarışıq variant da tətbiq edilir. Əksər web-forumlar qeydiyyatdan keçmiş istifadəçilərə fərdi ünsiyyət qurmağa imkan verən şəxsi məlumatlar sisteminə malik olurlar.

Qeydiyyat zamanı iştirakçılar öz profillərini yarada bilərlər. İştirakçı profili – onun haqqında məlumatları özündə əks etdirən web-səhifədir. Burada iştirakçı özü haqqında məlumat verə, şəklini daxil edə, onun məlumatlarına əlavə olunan imza yarada bilər. İmza mətn və ya qrafik (təsvir) şəklində ola bilər.

Beləliklə, web-forumlarda bölmələr, mövzular, məlumatlar, onlara cavablar, eləcə də mövzuların müəllifləri və iştirakçılar (qonaqlar) haqqında məlumatlar (onların profilləri) toplanır.

İnternet mağaza (ticarət) – internetdə malların alqı-satqısını həyata keçirən web-sayt olub istifadəçilərə saytda təqdim olunan malların alınması üçün sifariş tərtib etməyə, ödəmənin və ünvan çatdırmanın formasını seçməyə imkan verir. İnternet mağazalar müxtəlif mağazalar, ticarət və kommersiya təşkilatları, yayım və satış mərkəzləri və s. tərəfindən təqdim olunan məhsullar (qida, geyim, kitab, mebel, texnika, proqram məhsulları və s.) haqqında geniş məlumatları özündə saxlayır.

İnternet auksion – internet vasitəsilə həyata keçirilən auksiondur. İnternet auksion qiymətlər kütləvi və açıq ticarət zamanı sorğu və təkliflərə əsasən qoyulan internet ticarət sahəsidir.

Push texnologiya – istifadəçinin müəyyən etdiyi parametrlərə uyğun olaraq server tərəfindən informasiyanın ona təqdim olunması (göndərilməsi) yolu ilə web-saytların kontentinin yayılması texnologiyasıdır. Bir sıra ixtisaslaşmış web-serverlərdə müəyyən mövzular üzrə informasiya kanallarına yazılış təşkil olunur. Serverin mövzuları, server tərəfindən baxılan problemlər və ya təqdim olunan informasiya ehtiyatları ilə maraqlanan istifadəçilər bu serverdə qeydiyyatdan keçə bilirlər. Bundan sonra, server onun informasiya sistemində daxil olan bütün xəbərləri və yeni məlumatları müntəzəm olaraq həmin istifadəçinin ünvanına göndərir.

Baner sistemləri – saytlar arasında banerlərin mübadiləsini həyata keçirən web-saytlardır. **Baner** – üzərində basmaqla reklam səhifəsinə keçidi təmin edən qrafik təsvirlər və ya mətn blokudur. Banerlər web-saytlarda kommersiya və ya qeyri-kommersiya əsasında yerləşdirilir.

Banerlər istifadəçi tərəfindən açılan web-səhifələrdə avtomatik şəkildə peyda olur. Adətən, banerlərin əsas səhifənin məzmununa heç bir aidiyyəti olmur, lakin burada qoyulmuş hiperkeçidlər istifadəçini banerdə reklam olunan informasiyanın web-səhifəsinə və ya reklamçının serverinə aparıb çıxarır. Baner əlaqələri olan saytlar toplusunu **baner şəbəkələri** adlandırırlar.

İnformasiya-axtarış sistemləri – internetdə informasiya axtarışını həyata keçirməyə imkan verən web-interfeysli proqram-texniki kompleksidir.

2.3.2. E-mail – elektron poçtu

Elektron poçtu (“*e-mail*” və ya “*email*” – ingiliscə “electronic mail” sözlərindən) – paylanmış, o cümlədən

global (İnternet) kompüter şəbəkəsində elektron məlumatların göndərilməsi və alınması texnologiyası və xidmətidir. Elektron poçtu internetin ilk informasiya xidmətidir. O, yüksəksürətli və keyfiyyətli rabitə xətlərinin olmasını mütləq şəkildə tələb etmir.

Elektron poçtun yaranması tarixi kimi 1965-ci ili qeyd edirlər. Həmin ildə Massaçuset Texnoloji Universitetinin əməkdaşları Noel Morris və Tom Van Vlek IBM 7090/7094 kompüterində quraşdırılmış CTSS əməliyyat sistemi üçün MAIL proqramını yazmışdılar.

Elektron poçtu adı poçta uyğun olaraq məlumatların şəbəkənin (İnternetin) bir nöqtəsindən (qovşağından) digər nöqtəsinə (qovşağına) ötürülməsini təmin edir. Elektron poçtu vasitəsilə ötürülən məlumatlar “elektron məktub” və ya sadəcə “məktub” adlanır.

Lakin elektron poçtun adı poçt xidmətindən bir sıra üstünlükləri vardır. Elektron poçtu vaxtdan asılı olmur və ötürülən kimi ünvana çatır. Elektron məktuba mətnlə yanaşı qrafik, audio-video məlumatlar, hətta ikilik kodda fayllar (proqramlar) da əlavə edilə bilər. Elektron məktublar eyni zamanda bir neçə ünvana göndərilə bilər. Daxil olan elektron məktub başqa ünvana ötürülə bilər.

İnternetin istifadəçisi olan hər bir şəxs internetdə olan poçt serverlərindən birində (çox vaxt provayderə məxsus olan poçt serverində) özü üçün poçt qutusu açsın və ya açdıra bilər. İstifadəçinin göndərdiyi və aldığı bütün məktublar onun özü tərəfindən pozulmayanadək həmin poçt qutusunda saxlanılır.

Elektron poçtunun reallaşdırılması üçün SMTP, POP3, IMAP, MIME və digər protokollardan istifadə olunur.

Qeyd edildiyi kimi, elektron məktub, bir qayda olaraq, adı mətndən ibarət olur. Məktuba poçt proqramları tərə-

findən müəyyən xidməti məlumatlar əlavə edilir. Məktuba, həmçinin, onu göndərən istifadəçi tərəfindən qrafik, audio-video fayllar, şəkillər, Excel cədvəllər və s. qoşula (attachment) bilər.

Müasir dövrdə şəbəkədə ayrı-ayrı şəxslərin və təşkilatların identifikasiyası (tanınması) üçün onların elektron poçt ünvanlarından istifadə olunur. Bu ünvanlar informasiya-axtarış sistemlərində axtarış maşınları tərəfindən sərbəst olaraq indeksləşdirilə və sonradan informasiyanın axtarılması prosesində istifadə oluna bilər. Alta Vista sisteminin hesabatında göstərilir ki, axtarış sistemi tərəfindən indeksləşdirilmiş hər üç sənəddən ikisində elektron poçt ünvanlarına rast gəlinir.

Son statistik məlumatlara görə dünyada e-mail istifadəçilərinin sayı 3 milyarda, müştərilərin elektron poçtlarının (ünvanlarının) sayı isə 5 milyarda yaxındır. Statistik məlumatlara əsasən, hazırda internet istifadəçilərinin 87%-i elektron poçt xidmətindən istifadə edir.

Adətən, ayrı-ayrı şəxslər və təşkilatlar özlərinin veb-səhifələrinə “mailto” əmrinin köməyi ilə elektron poçt ünvanlarına hiperistinadları daxil edirlər. Zərurət yarandıqda istifadəçi belə istinadın üzərinə basır və bu zaman açılan elektron poçtunun müştəri proqramı vasitəsilə müvafiq ünvana məktub göndərir. Bundan əlavə, bir sıra interaktiv (online) internet xidmətlərində, o cümlədən sosial şəbəkələrdə, telekonfrans sistemlərində işləmək üçün istifadəçinin elektron poçt ünvanı zəruri rekvizit kimi tələb olunur.

Xüsusi olaraq adamların və təşkilatların axtarılması üçün nəzərdə tutulmuş informasiya-axtarış sistemlərində onların veb-fəzadan əldə edilməsi mümkün olan elektron poçt ünvanları toplanır. Lakin bu zaman nəzərə alınmalıdır ki, informasiya sistemlərində, xüsusən də pulsuz xid-

mət göstərən poçt serverlərində və digər xidmətlərdə istifadəçilər elektron poçt ünvanı əvəzinə öz təxəllüslerini (psevdonimlərini) istifadə etməklə qeydiyyatdan keçə və ya elektron poçt ünvanlarını dəqiq göstərməyə bilirlər.

Məlumatların göndərilməsi siyahıları – məktub əlaqəsi olan şəxslərə məlumatların kütləvi göndərilməsi məqsədi ilə yaradılan elektron poçt ünvanları siyahısıdır.

Elektron poçtun göndərilməsi siyahıları – elektron poçtu istifadəçilərinin kütləvi məlumatlandırılması, qruplar şəklində ünsiyyəti, reklam üçün istifadə olunur və elektron poçtun əvvəlcədən tərtib edilmiş siyahıya uyğun ünvanlar qrupuna avtomatik göndərilməsini təmin edir.

2.3.3. FTP – faylların ötürülməsi protokolu

FTP (ingiliscə File Transfer Protocol – faylların ötürülməsi protokolu) – TCP şəbəkələri (o cümlədən internet) vasitəsilə faylların ötürülməsi üçün nəzərdə tutulmuş standart protokoldur.

FTP faylların saxlandığı və istifadə olduğu nəhəng fayl arxivləri – informasiya saxlancısıdır. Web-texnologiyanın meydana gəlməsi ilə əlaqədar FTP xidməti öz əhəmiyyətini itirməmişdir və indiyə qədər istifadəçilər tərəfindən stabil olaraq istifadə edilir. Belə ki, hələ də FTP serverlərdə olduqca çox böyük həcmdə qiymətli informasiya ehtiyatları, proqram məhsulları, xüsusi məlumat bazaları və s. toplanmışdır.

Hazırda internetdə on minlərlə FTP serverlər fəaliyyət göstərir ki, onlarda yüz milyonlarla müxtəlif növ fayllar – proqramlar, qurğuların drayverləri, qrafik, audio-video fayllar və s. saxlanılır. İnternet istifadəçilərinin 42%-i FTP xidmətindən istifadə edirlər.

Web-texnologiyanın inkişafı nəticəsində, FTP xidmətinin hipermətn variantları yaradılmışdır. Lakin faylların alınması və ötürülməsinin sadəliyi, fayl sistemlərində işləməyin və naviqasiyanın rahatlığı, tam həcmdə axtarışın təşkilinin mümkünlüyü səbəbindən bu xidmətin web-fəzaya keçirilməsi hələ tam reallaşdırılmamışdır.

Eyni zamanda, internet şəbəkənin FTP arxivlərinin informasiya ehtiyatları üzrə axtarışın təşkili üçün axtarış sistemləri yaradılmışdır. Belə sistemlərdən biri olan Archie qlobal axtarış sistemidir və ona web-xidmət vasitəsilə daxil olmaq mümkündür. Archie sisteminə qoşulmaq imkanı verən web-serverlərdən birinin ünvanı belədir: <http://ftpsearch.com>.

FTP arxivlərin informasiya ehtiyatları üzrə axtarışın təşkili üçün, həmçinin, regional axtarış sistemləri də fəaliyyəti göstərir. Rusiya FTP ehtiyatları üzrə yaradılmış müvafiq axtarış sistemi aşağıdakı ünvana malikdir: <http://ftpsearch.ru>.

Bu cür axtarış sistemlərində informasiya axtarışı faylların və kataloqların adlarına görə aparılır.

2.3.4. Gopher – iyerarxik menyu

Gopher – internetdə sənədlərə giriş, onların paylanmış axtarışı və ötürülməsi üçün nəzərdə tutulmuş şəbəkə protokolidir. Gopher protokolu 1991-ci ildə Minnesot Universitetində işlənib hazırlanmış və 1993-cü ilədək internetdə geniş istifadə olunmuşdur.

Gopher sistemi iyerarxik menyu (ağacvari struktur) şəklində təqdim olunan və əsasən, mətn tipli məlumatları özündə saxlayan informasiya sistemidir. Onun ilkin kata-

loqu iyerarxiyanın başlanğıcı (ağacın kötüyü), qalanları isə menyu və ya altmenyulardan ibarətdir. Gopher menyusunun hər bir bəndi ya altmenyuuya, ya da konkret informasiyaya keçidi təmin edir.

Hazırda, demək olar ki, Gopher sisteminin bütün informasiya ehtiyatları hipermətn texnologiyasının köməyi ilə web mühitinə keçirilmişdir.

2.3.5. Telnet – uzaq məsafədən idarəetmə və sistemə daxilolma

Telnet (ingiliscə Terminal NETwork sözlərindəndir) – şəbəkə vasitəsilə mətn tipli interfeysin reallaşdırılması üçün nəzərdə tutulmuş şəbəkə protokolidir. 1983-cü ildə reallaşdırılmışdır. Telnetin təyinatı kifayət qədər ümumi, iki istiqamətli, bayt-yönümlü səkkizbitlik rabitə vasitəsinin təqdim edilməsi ilə bağlıdır. Onun əsas funksiyası terminal qurğularına (“terminal-terminal” rabitəsi) və terminal proseslərinə (paylanmış hesablama rabitəsi) bir-biri ilə qarşılıqlı işləmək imkanı yaratmaqdan ibarətdir.

Telnet xidməti telnet protokolidə əsasında qurulan və uzaq məsafədən kompüterlə qarşılıqlı əlaqəni təmin edən xidmətdir. Telnet vasitəsilə əlaqə yaratdıqdan sonra istifadəçi uzaq məsafədə olan kompüterlə öz kompüterini kimi işləyir. İstifadəçinin kompüterini telnet vasitəsilə qoşulan kompüterin terminalına çevrilir. İstifadəçi həmin kompüterdə qurulmuş informasiya təhlükəsizliyi siyasətinə uyğun olaraq icazə verilən bütün ehtiyatlara giriş əldə edir.

İnternet şəbəkəsində WWW, FTP, e-mail və digər xidmətlər daha geniş yayıldığından *telnet* protokolidə (eləcə də xidməti) çox az istifadə olunur. Lakin buna baxmayaraq,

yalnız *telnet* protokolu vasitəsilə giriş imkanı olan, əhəmiyyətli və faydalı məlumatları özündə saxlayan unikal informasiya sistemləri indiyə qədər də fəaliyyət göstərir. Bu növ sistemlərə lokal şəbəkədə istifadə üçün nəzərdə tutulmuş informasiya sistemləri, o cümlədən kitabxanaların elektron kataloqları, dövlət təşkilatlarının avtomatlaşdırılmış informasiya sistemləri (kadrların uçotu, müəssisələrin fəaliyyətinin və istehsalatının idarə olunması), eləcə də internetin elektron elan lövhələri (BBS) və s. aid edilir.

2.3.6. BBS – elektron elan lövhələri

BBS (ingiliscə Bulletin Board System – elektron elan lövhəsi) xüsusi proqram təminatı olan, modemlə telnet protokolu və ya internetdə IP protokolu vasitəsilə qoşulmağa imkan verən xidmətdir. Bu xidmət ayrıca kompüterin üzərində reallaşdırıla bilər. BBS ünsiyyət və faylların mübadiləsi üçün nəzərdə tutulmuşdur. BBS ilk dövrlərdə paylanmış kompüter şəbəkələri geniş yayılmadığı dövrdə telefon xətti və modem vasitəsilə nömrə yığaraq (Dial-Up) qoşulmaqla həyata keçirilirdi. Lakin internet yayıldıqca BBS adını özündə saxlayan adi elektron elan lövhələrinə və reklam qəzetlərinə oxşar saytlar meydana gəlməyə başladı.

Bu gün elektron elan lövhələri kommersiya və ya qeyri-kommersiya əsaslı elanların və reklamların yerləşdirildiyi saytdır. Bu saytlarda bir sıra mövzu bölmələri yaradılır, reklam və elanlar məzmunlarına uyğun olaraq müvafiq bölmələrə yerləşdirilir. Bir qayda olaraq, elektron elan lövhələri qeyri-kommersiya xarakterli olur. Əksər elektron elan lövhələri mal və xidmətlərin alqı-satqısı haqqında elanların yer-

ləşdirilməsi üçün nəzərdə tutulur, lakin başqa xidmətlər təklif edən elan lövhələri də mövcuddur.

Müasir elektron elan lövhələri mal və xidmətlərin ətraflı təsvirini, şəkillərini, dəqiq qiymətini, yerləşdiyi yeri və çatdırılma şərtlərini, əlaqə kanallarını (ünvan, telefon nömrələri, e-mail ünvanı və s.) əlavə etməyə imkan verir. Elanın başlığında və mətnində açar sözlərin istifadəsi həmin elanların axtarış sistemləri tərəfindən emal olunmasını və axtarış bazasına düşməsini təmin edir.

2.3.7. Usenet – telekonfrans sistemi

Usenet sistemi beynəlxalq aləmdə kompüterlər arasında xəbərlərin yayılması üçün işlənilib hazırlanmışdır. Sonralar Usenet internetin xidmətinə çevrilərək, onun bütün məlumatlarının yayılmasını təmin edir. Usenet serverlərində müəyyən mövzular üzrə telekonfransların keçirilməsi üçün imkanlar və vasitələr reallaşdırılır.

Telekonfranslar iyerarxiya prinsipi əsasında təşkil olunurlar. Usenet sisteminin ən yuxarı səviyyəsi üçün yeddi əsas rubrika (məsələn, “comp”, “sci”, “news”, “soc”, “talk” və s.) fəaliyyət göstərir. Bunlardan əlavə, xüsusi və regional rubrikalar yaradıla bilər. Bu rubrikaların hər birində yüzlərlə bölmələr – Usenet qrupları mövcuddur. Onları çox vaxt xəbər qrupları adlandırırlar. Hər bir qrup unikal ada malik olur. Bu ad global Usenet sistemi üçün açar söz rolunu oynayır.

Hal-hazırda Usenet sisteminin xəbər qruplarının sayı on minlərlə ölçülür. Təbii ki, telekonfrans xidmətinin bütün qruplarına xidmətin bir server tərəfindən həyata keçirilməsi mümkün deyil, ona görə də bu xidmətlər internetə qoşulmuş müxtəlif serverlər üzrə paylanırlar.

İstifadəçilər Usenet sistemində mövcud olan mövzular üzrə müvafiq xəbərlər qrupuna yazıla, yəni orada qeydiyyatdan keçə bilirlər. Usenet sistemində qeydiyyatdan keçmiş istənilən istifadəçi müvafiq xəbərlər qrupuna daxil olan mövzular üzrə məlumatlardan istifadə etməklə yanaşı, bu mövzular üzrə öz informasiyasını da orada yerləşdirə bilər. İnformasiya daxil edilən kimi həmin qrupun bütün istifadəçilərinə ötürülür. Usenet sistemində işləyərkən əsas çətinlik lazım olan mövzunu, yəni xəbərlər qrupunu saxlayan serverin tapılması ilə bağlı olur.

Bir qrupa daxil olan bütün istifadəçilər ümumi marağa malik olduğuna görə imkan daxilində bir-birlərinə kömək etməyə çalışırlar. Hər hansı məlumat barəsində sorğu alan və ya həmin informasiyaya malik olan istənilən istifadəçi onu bilavasitə ehtiyacı olan istifadəçiyə göndərə bilər. Usenet sistemi, əsasən, hər hansı konkret məsələ üzrə məlumatın, şəxsi və ya qeyri-rəsmi informasiyanın alınması üçün çox əlverişlidir.

2.3.8. İnternetin interaktiv informasiya xidmətləri

İnternetin interaktiv xidmətləri internetdə xüsusi serverlərin köməyi ilə insanlar arasında real zaman kəsiyində bilavasitə ünsiyyətin qurulmasını həyata keçirən texnologiyalardır. İnformasiya mübadiləsi xüsusi müştəri program təminatlarının köməyi ilə mətnlərin yazılması, qrafik təsvirlərin çəkilməsi, multimediyalı səs və video imkanları olan web-kameraların istifadəsi yolu ilə reallaşdırılır.

İnternetdə interaktiv informasiya mübadiləsini həyata keçirən xidmətlər qismində ICQ (Internet Relay Chat), Skype, NetMeeting, MUD (Multi-User Dungeon), MOO

(MUD Object-Oriented) və digər xüsusi serverlərin köməyi ilə real zaman kəsiyində (interaktiv) iki və daha çox internet istifadəçisi arasında informasiya mübadiləsinin həyata keçirilməsi üçün nəzərdə tutulmuşdur.

Bu sistemin xidmətlərindən istifadə etmək istəyən istifadəçilər öz kompüterində müvafiq müştəri programını quraşdırmalı, uyğun serverə qoşularaq qeydiyyatdan keçməlidirlər. Qeydiyyatdan keçmiş hər bir istifadəçiyə unikal ad, yəni identifikator nömrəsi verilir. Bundan sonra, həmin seansın gedişində istifadəçi bu identifikatorla tanınır və yazışmalar aparır.

Yuxarıda sadalanan interaktiv xidmətlərlə yanaşı, son illərdə internet-telefon, internet-televiziya, internet-radio və s. kimi yeni texnologiyalar meydana gəlmişdir.

İnternet-telefon – internet və ya istənilən IP-şəbəkə vasitəsilə səs siqnallarının ötürülməsini təmin edən kommunikasiya protokol, üsul və vasitələrini özündə birləşdirən texnologiyadır. İnternet-telefon termini əvəzinə IP-telefon və ya VoIP (Voice over IP) terminlərindən istifadə olunur.

İnternet-televiziya – genişzolaqlı qoşulma yolu ilə internet vasitəsilə televiziya siqnallarının ikitərəfli rəqəmsal ötürülməsinə əsaslanan texnologiyadır. Ona bəzən “online TV” deyirlər. İnternet-televiziya dünyanın istənilən nöqtəsindən internet vasitəsilə televiziya verilişlərinə interaktiv (online) rejimdə baxmağa imkan verir. İnternet-televiziya yenicə inkişaf etməyə başlayan və çox geniş miqyas alan texnologiyadır. Onun istifadəçiləri yaxın vaxtlaradək mümkün olmayan televiziya kontentini tətbiq etmək üçün geniş imkanlar əldə edirlər.

İnternet-radio – internet vasitəsilə audioverilənlər axınının ötürülməsi texnologiyasıdır. İnternet-radio dedikdə

radioverilişin yayılması üçün internetdə verilənlər axınının ötürülməsi texnologiyasını istifadə edən radiostansiya başa düşülür. Bu zaman səs verilənləri ilə yanaşı qəbul edicidə əks olunması üçün nəzərdə tutulan mətn informasiyası (radiostansiya və musiqi haqqında) da göndərilir. Burada radiostansiya qismində adi audio-pleyer proqramı və ya xüsusi proqram təminatı, eləcə də analoq signalını rəqəmsal verilənlərə çevirən aparat qurğusu çıxış edə bilər.

2.4. İnternet şəbəkəsinin informasiya ehtiyatlarının xüsusiyyətləri

Qeyd olunduğu kimi, internetin informasiya xidmətlərinin yaradılması texnologiyası istənilən informasiyanın şəbəkədə yerləşdirilməsi, saxlanması, alınması və ötürülməsi üçün çox rahat və əlverişli imkanları təqdim edir. Lakin, digər tərəfdən internetdə informasiya ehtiyatlarının, massivlərinin, arxivlərinin həcmnin və sayının kəskin artması və s. ilə əlaqədar lazım olan məlumatların tapılması getdikcə daha çətin və problemlə məsələyə çevrilmişdir.

Lazım olan məlumatların əldə olunması üçün istifadəçilər hər dəfə müxtəlif informasiya sistemləri ilə əlaqə yaratmalı, ardıcıl olaraq bir informasiya ehtiyatından digərinə keçməli, beləliklə də, bütün şəbəkəni "gəzməli" (araşdırmalı) olurlar. Bu zaman istifadəçilər informasiya axtarışına külli miqdarda vaxt, eləcə də müəyyən informasiya, şəbəkə və maddi resurslar sərf etmək məcburiyyətində qalırlar. Nəicədə şəbəkəni izafi yükləmiş olurlar.

Belə sistemlər zəruri informasiyanın axtarılmasına və seçilib çıxarılmasına əhəmiyyətli kömək edə, bəzən isə (hətta çox vaxt) əksinə, əbəs yerə vaxt itkisinə səbəb ola

və sonsuz sayda web-qovşaqlar labirintində istifadəçinin "azmasına" gətirib çıxara bilər.

Qeyd etmək lazımdır ki, internetin informasiya fəzası öz təbiətinə görə paylanmış informasiya sistemidir. Lakin lokal və ya korporativ şəbəkələrin məlumat bazalarında (hətta məlumat bazasının müstəqil olması mümkündür) yerləşən və şəbəkə administratorunun nəzarəti altında olan ənənəvi informasiya sistemlərindən fərqli olaraq, internetin informasiya fəzası bir sıra spesifik xüsusiyyətlərə malikdir. Bu xüsusiyyətlər internetdə reallaşdırılan texnologiyalardan irəli gələn, üstünlük və çatışmazlıqlara səbəb olan amillərlə xarakterizə olunur:

- ***İnformasiya fəzasının ölçüsü.*** İnternetin yüz milyonlarla saytlarında və qovşaqlarında yerləşən informasiyanın həcmi olduqca böyükdür. İnformasiya ehtiyatları internetin müxtəlif qovşaqları üzrə paylanır, yəni onların bir hissəsi bir serverdə, digər hissəsi isə başqa serverdə yerləşmiş olur. Ona görə də informasiya massivinin bütün ehtiyatlarının axtarış sistemləri tərəfindən tamamilə əhatə olunmasında problemlər mövcud olur.
- ***Xaotiklik.*** İnternetin informasiya ehtiyatlarının yaradılması, toplanması və saxlanılmasında hər hansı vahid qanunauyğunluq olmur, onlar qovşaq və serverlər üzrə paylanmasına görə xaotik olurlar. Başqa sözlə, informasiya ehtiyatları dünyanın müxtəlif yerlərində olan müxtəlif şəbəkə qovşaqları üzrə əhəmiyyətli dərəcədə dağınıq formada səpələnirlər və nəticədə şəbəkə xaosu yaranır.
- ***Sistemsizlik.*** Yeni informasiyanın yaradılmasının, köhnə informasiyanın dəyişdirilməsinin və məhv edilməsinin çox asan olması səbəbindən internet şəbəkəsinə

aidiyyəti olub-olmamasından asılı olmayaraq, hər bir arzu edən istifadəçi özünün şəxsi səhifəsini yaratmaq, bu səhifəyə istənilən informasiyanı daxil etmək, onu internetdə yerləşdirmək və ya oradan çıxarmaq imkənina malikdir. Beləliklə, internet şəbəkəsində istənilən informasiya təsadüfi şəkildə meydana gələ və yox ola bilər. Ümumiyyətlə, informasiya ehtiyatlarının, o cümlədən web-səhifələrin və ya saytların informasiya təminatının təşkili prinsipləri serverlərin sahiblərinin arzusundan asılı olaraq seçilir və həyata keçirilir. Məhz bu baxımdan web-fəzada informasiya ehtiyatlarının sistemliliyini təmin etmək mümkün olmur.

- **Natamamlıq, izafilik, ziddiyyət, doğruluq, adekvatlıq.** Yuxarıda göstərilən səbəbdən, habelə informasiya ehtiyatları internetdə müstəqil şəkildə, bir-birindən asılı olmadan, müxtəlif vaxtlarda və yerlərdə yaradıldığından, təbii ki, onların bir-biri ilə uzlaşdırılmaması problemi meydana çıxır. Bu baxımdan web-fəzada informasiya ehtiyatlarında natamamlıq və ya izafilik (həddən artıq bolluq), doğru olmamaq və qeyri-adekvatlıq, eləcə də onlar arasında qarşılıqlı ziddiyyət hallarının olması mümkündür.
- **Müxtəlif dillər və kodlar.** İnformasiya ehtiyatları ayrı-ayrı adamlar tərəfindən müxtəlif yerlərdə (şəhərlərdə və ölkələrdə) müxtəlif məqsədlər üçün bir-birindən fərqlənən kompüterlərdə və sistemlərdə yaradılması səbəbindən informasiya ehtiyatlarının dillərinin və kodlarının (xüsusən də milli informasiya ehtiyatları üçün) müxtəlifliyi ilə bağlı əlavə problemlər yaranır.
- **Terminologiyanın müxtəlifliyi.** Müəlliflər tərəfindən istifadə olunan terminologiya informasiya ehtiyatlarının hazırlanması tarixindən, yaranma şəraitin-

dən və təyinatından, hazırlanmasına elmi-metodiki yanaşmadan, müəlliflərin fərdi qabiliyyətlərindən və bacarıqlarından və s. asılı olaraq dəyişir. Buna görə də informasiya axtarışı və sənədlərin tematik profillər üzrə təsnifatı zamanı problemlər yaranır və sorğu lüğətləri, tezaurus, assosiativ sözlər kimi əlavə aparatın tətbiq edilməsi zərurətini yaradır.

- ***İnformasiyanın qiymətliliyi və həyat dövrü (yaşama qabiliyyəti).*** İnternetə daxil edilən və müəyyən qiymətliliyi ilə seçilən informasiya ehtiyatları müəyyən müddətdən sonra öz əhəmiyyətini itirə və ya köhnələ bilər. Bəzən internetə daxil edilən informasiya ehtiyatları onların müəllifindən başqa heç kəs üçün maraq və əhəmiyyət kəsb etmir. Bəzən isə informasiya ehtiyatları internetdə peyda olduğu anda artıq qeyri-aktual və köhnəlmiş olur. Eyni zamanda qeyd olunmalıdır ki, web-səhifələr və serverlər ayrı-ayrı adamlar və təşkilatlar tərəfindən yaradılsa da, sonradan onlara xidmət göstərilir, yəni saytda və səhifədə olan məlumatlar yenilənmir, tamamlanmır, dəyişdirilmir və s. Başqa sözlə, onlar atılmış "informasiya zibillərinə" çevrilirlər.
- ***Hiperistinadların istifadəsinin üstün və çatışmayan cəhətləri.*** Hipermetn texnologiyasının tətbiqi bir tərəfdən tematika üzrə yaxın olan bir serverdən digərinə keçidi təmin etməklə işi xeyli asanlaşdırır, digər tərəfdən o, düzgün olmayan istiqamətə («dalana») yönəldə və ya tamamilə lazım olmayan yerə göstərməklə çıxılmaz vəziyyətə gətirib çıxara bilər. Belə ki, hiperistinadlar müəlliflərin subyektiv fikirlərinə əsasən qoyulur. Ona görə də onlar lazımsız, heç bir qiymətə malik olmayan, çoxdan köhnəlmiş, dəyişdi-

rılmış, pozulmuş və başqa yerə köçürülmüş səhifələrə və ya saytlara istinad edə bilirlər

İnternetdə olan informasiya sistemləri arasındakı fərqləri nümayiş etdirmək məqsədilə informasiya ehtiyatlarının aşağıdakı növlərini sadalamaq kifayətdir:

- *elektron nəşrlər* – dövri elektron qəzetlər və jurnallar, dissertasiyalar, avtoreferatlar, monoqrafiyalar, icmallar, bülletenlər, konfrans, seminar və simpozium materialları və s.
- *elektron kitabxanalar və kataloqlar* – ənənəvi kitabxanalara analoji olan elektron kitabxana fondları, eləcə də hər iki növ (həm ənənəvi, həm də elektron) kitabxana fondlarının elektron kataloqları və s.
- *İnternetin informasiya massivi* – web, FTP arxivlər, Gopher ehtiyatları, Usenet xəbərlər sistemi və digər informasiya sistemləri.
- *xüsusi təyinatlı informasiya ehtiyatları* – xidməti məqsədlərlə kompüterə daxil edilmiş məlumatlar, hesabatların mətnləri, elmi tədqiqatların nəticələri, məqalələr, müxtəlif sənədlər və s.

2.5. İnformasiya ehtiyatlarının məzmunlarının təhlilinə əsaslanan empirik paylanma qanunları

İnsanlar tərəfindən hazırlanmış informasiya ehtiyatlarının (sənədlərin) məzmununun statistik tədqiqi nəticəsində müəyyən edilmişdir ki, mətnlərində sözlərin istifadəsi və paylanması ilə bağlı olan, müəyyən bibliometrik prinsipləri özündə ehtiva edən bir sıra qanunlar mövcuddur. Bu qanunlara Şennon, Zipfin I və II, Bredford, Pareto, Parkinson qanunlarını aid etmək olar.

İnformasiya axtarışı baxımından Zipf və Bredford qanunları bilavasitə maraq kəsb edir.

Tədqiqatlar göstərir ki, insanlar tərəfindən yaradılmış mətnlər bioloji və psixoloji amillərdən asılı olaraq müəyyən xüsusiyyətlərə malik olurlar. Belə ki, onlar mətn hazırlayarkən özlərindən asılı olmadan müəyyən çərçivədən kənara çıxma bilmirlər.

Zipf müəyyən etmişdir ki, mətnlərdə böyük uzunluğa malik olan sözlər kiçik uzunluğa malik olan sözlərə nisbətən daha az istifadə edilir, eləcə də terminlərin mətnlərdə rast gəlinmə tezliyi ilə sənədlərin məzmununa uyğunluğu arasında asılılıq mövcuddur. Bu postulat əsasında Zipf iki universal mülahizə irəli sürmüşdür.

İnformasiya ehtiyatlarının məzmunlarının (mətnlərinin) statistik təhlili üçün irəli sürülmüş Zipf qanunları mətnlərdə rast gəlinən sözlər çoxluğundan termin kimi qəbul oluna biləcək daha yığcam altçoxluğu müəyyən etməyə, ümumiyyətlə, mətnlərin xarakteristikalarını aydınlaşdırmağa imkan verir.

Bredford isə sənədlərin (məqalələrin, web-səhifələrin və s.) məzmunları ilə onların müəyyən mənbələr (jurnallar, web-saytlar və s.) üzrə paylaşılması arasında qanunauyğunluq aşkar etmişdir.

2.5.1. Zipfin I qanunu – “ranq-tezlik” asılılığı

Zipfin I qanunun mahiyyəti aşağıdakıdan ibarətdir. Sənədin mətninin statistik təhlili aparılır. Mətnə rast gəlinən bütün sözlərin sənədə daxilolma tezlikləri (sözlərin mətnə rastgəlinmələrinin sayı) müəyyən edilir. Aydındır ki, bir sıra sözlər eyni tezliyə malik ola bilərlər. Sözlər mətnə rastgəlinmə

tezliklərinə görə qruplaşdırılır. Tezlikləri eyni olan sözlərin hamısı bir qrupda birləşdirilir. Bütün qruplar onlara daxil olan sözlərin tezliklərinin qiymətinə görə azalan sıra ilə düzülür və ardıcıl olaraq nömrələnirlər. Qrupların sıra nömrəsi onların ranqı adlanır. Belə ki, ən çox rast gəlinən sözlər qrupunun ranqı 1, ondan sonrakı 2 və s. olur.

Zipf sübut etmişdir ki, sənədə daxil olan sözlərin ranqları ilə onların rastgəlinmə tezlikləri arasında asılılıq mövcuddur.

Tutaq ki, $f_j - j$ sözünün sənədin mətninə daxilolma tezliyi, $f_d -$ sənədin mətnindəki sözlərin ümumi sayıdır. Onda sözün mətnədə rast gəlinməsi ehtimalı aşağıdakı kimi olur:

$$p_j = \frac{f_j}{f_d}.$$

Zipf “ranq-tezlik” asılılığını aşağıdakı şəkildə ifadə etmişdir: sözlərin mətnədə rastgəlinmə ehtimalı onların ranqları ilə düz mütənasibdir. Başqa sözlə, sözlərin mətnədə rast gəlinmə ehtimalının onların ranqına hasili nəticəsində alınan qiymət təxminən sabit qalır:

$$c \approx p_j \cdot r_j = \frac{f_j \cdot r_j}{f_d}.$$

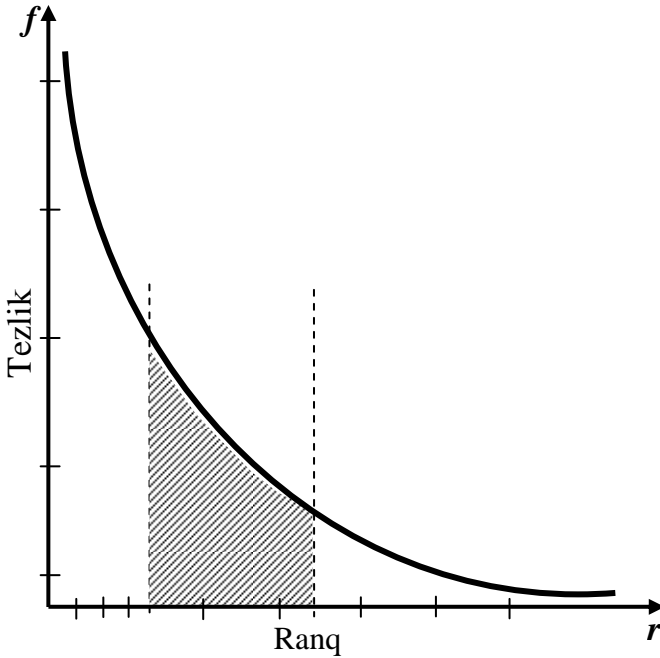
Burada $r_j - j$ sözünün ranqı, $c -$ empirik sabitdir. Bu sabit müxtəlif dillər üçün fərqli ola bilər, lakin onlar təxminən yaxın olur. Məsələn, bu sabit ingilis dili üçün 0.1, rus dili üçün 0.06-0.07, Azərbaycan dili üçün təxminən 0,09 qiymətlərini alır.

Zipfın I qanununa əsasən, sözlərin rastgəlinmə tezlikləri ilə ranqları arasındakı asılılıq hiperbola şəklində

olur və sənəd üçün daha çox məna kəsb edən, yəni termin kimi qəbul oluna biləcək sözlər qrafikin orta hissəsində yerləşir (şək.2.4).

Daha çox rast gəlinən (rastgəlinmə tezliyi daha böyük olan) sözlər əsasən əvəzlilər, bağlayıcı və s. kimi xidməti sözlərdir. Ən az rast gəlinən sözlər isə fellər və digər xidməti, eləcə də təsadüfi istifadə olunan sözlər olur. Ona görə də həddən artıq çox və həddən artıq az rast gəlinən sözlər, adətən, mətnin məzmununa yaxınlıq baxımından əhəmiyyət kəsb etmirlər. Açar sözlərə və terminlərə isə, əsasən, orta rastgəlmə tezliyinə malik olan sözlər aid olurlar.

Burada hər şey sözlərin rastgəlinmə tezliyi intervalının necə seçilməsindən asılı olur. Belə ki, bu interval çox



Şək.2.4. Zipfin I qanunu: “ranq-tezlik” asılılığı

geniş götürülsə, onda köməkçi və xidməti sözlər termin qismində sənədin axtarış obrazına düşə bilər ki, bu da sonradan axtarış prosesini çətinləşdirər. Əgər bu interval çox kiçik seçilərsə, onda əhəmiyyət kəsb edən terminlər diqqətdən kənar qala bilərlər. Praktika göstərir ki, bu interval əsasən rənglər siyahısının ortasında yerləşən bir zolaqdan ibarət olur.

Nümunəyə baxaq. Tutaq ki, mətndə “Azərbaycan”, “respublika”, “ölkə” və s. sözlər təxminən 30 dəfə ($f_1=30$), “müstəqil”, “bayraq”, “müdafiə” və s. sözlər 15 dəfə ($f_2=15$) dəfə, “əsgər”, “torpaq”, “sərhəd” və s. sözlər təxminən 10 dəfə ($f_3=10$) dəfə, “düşmən”, “müharibə” və s. sözlər isə 8 dəfə ($f_4=8$) dəfə rast gəlinmişdir. Onda $c=30 \cdot 1=15 \cdot 2=10 \cdot 3=8 \cdot 4=\dots \approx 30$ olar.

2.5.2. Zipfin II qanunu – “say-tezlik” asılılığı

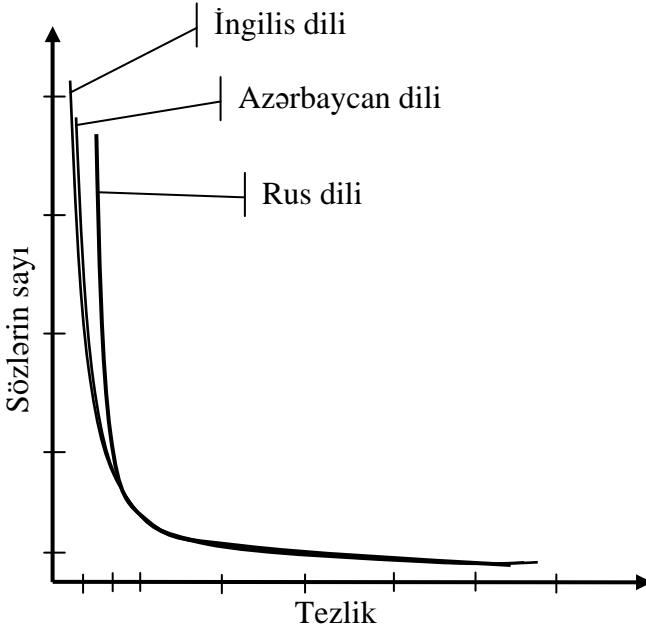
Zipfin I qanununda müxtəlif sözlərin mətne eyni tezliklə daxil olması amili diqqətdən kənar qalmışdır. Sözlərin mətnlərə daxilolmalarının statistik göstəriciləri üzərində təhlilini davam etdirən Zipf müəyyən etmişdir ki, sözlərin mətnlərə daxilolma tezlikləri ilə həmin tezliyə malik sözlərin sayı arasında da asılılıq mövcuddur.

Belə ki, insanlar tərəfindən yaradılmış bütün mətnlər üçün tezlik və həmin tezliyə malik sözlərin sayı arasındakı asılılıq hiperbola şəklidə olur. Bu qanuna uyğunluq bütün dillər üçün öz qüvvəsini saxlayır. Yalnız əyrinin meyilliyini müəyyən edən əmsal cüzi fərqlənə bilər (şək.2.5).

Qeyd edilməlidir ki, Zipfin qanunları universal olub bir çox sahələrdə doğruluğunu saxlayır. Məsələn, şəhərlərdə

yaşayan əhalinin sayı ilə həmin sayda əhalisi olan şəhərlərin sayı arasındakı asılılıq, varlı insanların (şirkətlərin) kapital-ları ilə həmin həcmdə kapitalla malik insanların (şirkətlərin) sayı arasındakı asılılıq və s. Zipf qanununa cavab verir.

Nümunəyə baxaq. Tutaq ki, “Azərbaycan”, “respublika”, “ölkə” və s. cəmi 20 söz hərəsi 30 dəfə, “müstəqil”, “bayraq”, “müdafiə” və s. cəmi 40 söz hərəsi 15 dəfə, “əsgər”, “torpaq”, “sərhəd” və s. cəmi 60 söz hərəsi 10 dəfə, “düşmən”, “müharibə” və s. cəmi 80 söz hərəsi 8 dəfə və s. rast gəlinmişdir. Onda $c=30 \cdot 20=15 \cdot 40=10 \cdot 60=8 \cdot 80=... \approx 600$ olar.



Şək.2.5. Zipfin II qanunu:
“sözlərin sayı – tezlik” asılılığı

2.5.3. Bredford qanunu

Elmi jurnalların məzmunlarının, o cümlədən hər hansı mövzu sahəsi üzrə məqalələrin elmi jurnallar üzrə paylanmasının tədqiqi zamanı Bredford elmi jurnalların sayı ilə bu və ya digər mövzu sahəsi üzrə həmin jurnallarda dərc olunmuş məqalələrin sayı arasında qanunauyğunluq müəyyən etmişdir.

Belə ki, əgər elmi jurnallar hər hansı mövzu sahəsi üzrə çap olunmuş məqalələrin sayına görə azalan sıra ilə düzülərsə, onda onları elə zonalara bölmək olar ki, hər bir zonda həmin mövzu sahəsi üzrə məqalələrin sayı eyni olsun.

Tutaq ki, $N_1, N_2, N_3, N_4, \dots$ – uyğun olaraq 1-ci, 2-ci, 3-cü, 4-cü və s. zonalarda olan jurnalların sayı, k – bir zonadakı jurnalların sayının növbəti zonadakı jurnalların sayına nisbətini göstərən kəmiyyətdir. Onda

$$N_1 : N_2 : N_3 : N_4 : \dots = 1 : k : k^2 : k^3 : \dots$$

olar.

Bredford qanunu internetin saytlarının və saytlarda dərc olunan web-səhifələrin sayı arasındakı münasibət üçün də doğrudur.

Bu qanunauyğunluq informasiya fəzasında (o cümlədən internetin web-fəzasında) çox böyük ehtimalla məhz istifadəçini maraqlandıran mövzu sahəsinə uyğun olan məqalələri (web-səhifələri) mövzu özündə saxlayan nəşrləri (web-saytları) seçməyə imkan verir.

Nümunəyə baxaq. Tutaq ki, bizi maraqlandıran mövzu sahəsi üzrə məqalələrin sayına görə ilk on yeri tutan cəmi 10 jurnal (sayt) müəyyən edilmişdir. Əgər bu zonada məqalələrin (web-səhifələrin) cəmi sayı 30 olarsa, onda daha 30 məqalənin (web-səhifənin) tapılması üçün mövcud jurnalların sırasına əlavə $10 \cdot k$ sayda (əgər $k=5$ olarsa, 50) jurnal (sayt)

əlavə edilməlidir. Üçüncü mərhələdə növbəti 30 məqalə (web-səhifə) üçün siyahıya daha $10 \cdot k^2$ sayda (yəni 250) jurnal əlavə edilməlidir.

Göründüyü kimi, əvvəlcə 10 jurnalдан (saytdan) 30 məqalə (sənəd) tapılmışdırsa, ikinci mərhələdə yalnız 50 jurnaldan (saytdan) 30 məqalə, üçüncü mərhələdə isə 250 jurnaldan (saytdan) 30 məqalə tapılar.

Beləliklə, ilk 10 jurnalın (saytın) baxılması mövzu sahəsi üzrə əsas məqalələri əldə etməyə imkan verir.

2.5.4. Pareto qanunu

Pareto müəyyən etmişdir ki, tənəsübsüzlük (uyğunsuzluq) səbəblə (sərf olunan əməklə, kapital qoyuluşu ilə) nəticə (əldə olunan nəticə, qazanılan gəlir) arasındakı münasibətin əsas ayrılmaz xassəsidir.

Belə ki, mütəxəssislərin fikrinə görə, nəticənin əhəmiyyətli dərəcədə böyük hissəsi (təxminən 80%-i) səylərin çox kiçik hissəsinin (təxminən 20%-nin) hesabına əldə olunur. Səylərin əhəmiyyətli dərəcədə böyük hissəsi (təxminən 80%-i) nəticənin yalnız qalan az hissəsinin (təxminən 20%-nin) əldə olunması təmin edir.

Məsələn, hər hansı fəaliyyətə həsr olunmuş əməyin 20%-i əldə edilmiş nəticənin 80%-ni verir, qalan 80% əmək isə 20% nəticənin əldə olunmasına sərf olunur. Müştərilərin 20%-i gəlirin 80%-ni, qalan 80% müştəri isə yalnız 20% verir. Dünya əhalisinin 80%-i böyük şəhərlərin 20%-də toplanmışdır, əhalinin yalnız 20%-i dünya şəhərlərinin qalan 80%-i üzrə paylanmışdır.

Məhz 20 və 80 tərs mütənəsbliyəndən irəli gələrək, Pareto qanunu çox vaxt “20/80 prinsipi” adlandırılır. Lakin, nəzərə

alınmalıdır ki, burada 20 və 80 ədədləri dəqiq deyil, təxmini götürülür, yəni təxminən 20% və ya təxminən 80%.

Pareto qanunu müxtəlif sahələrə tətbiq olunur. Bu qanunauyğunluq internet ehtiyatları və xidmətləri üçün öz doğruluğunu saxlayır. Məsələn, konkret mövzu sahəsi üzrə informasiya ehtiyatlarının 80%-i yalnız 20% saytda toplanmışdır. İnternet istifadəçilərinin 80%-i yalnız 20% saytdan istifadə edirlər. İnternetdə həyata keçirilən informasiya axtarışının 80%-i yalnız 20% axtarış sistemləri tərəfindən həyata keçirilir.

2.6. İnternetdə istifadə olunan əsas fayl formatları

Məlumdur ki, internet saytlarda müxtəlif növ məlumatlara rast gəlmək mümkündür. Təbii ki, web-səhifələrdə bu məlumatlar ayrı-ayrı formatlara malik fayllar və ya verilənlər şəklində təqdim olunurlar. Web-səhifələrin istifadəçinin kompüterində normal əks olunması və sonradan onların istifadəsi üçün internetdə tətbiq olunan formatlar və bu formatları reallaşdıran proqram təminatları barədə ilkin biliklərin olması zəruridir. Belə ki, informasiya-axtarış sistemləri istifadəçinin sorğusuna uyğun məlumatları tapıb ona təqdim edir, lakin müvafiq format istifadəçiyə bəlli olmadıqda və kompüterə aidiyyəti proqram yüklənmədikdə həmin məlumatın istifadəsi mümkün olmur. Bununla əlaqədar olaraq, internetdə istifadə olunan əsas (ən çox rast gəlinən) fayl formatları barədə aşağıda məlumat verilir.

Arxiv faylları:

- **RAR** – Rar vasitəsilə sıxılmış fayl və ya fayllar qrupu;
- **ZIP** – Zip vasitəsilə sıxılmış fayl və ya fayllar qrupu;
- **ACE, ARJ, BZIP (.bz2), GZIP (.gz), LHA, TAR, ZOO** – MS DOS, MS Windows və Unix əməliyyat sistemləri mühitində reallaşdırılan arxiv faylları.

Video fayllar:

- **AVI** (Audio Video Interleave) – Windows platformasında dəstəklənən standart video format;
- **MPEG/MPG** (Motion Picture Experts Group) – Filmlərin və videokliplərin axınla ötürülməsi üçün standart MPEG standartı vasitəsilə sıxılmış format;
- **SWF (ShockWare Flash)** – Macromedia Flash proqramı vasitəsilə yaradılan video fayllar.

Qrafik faylların formatları:

- **AI** (Adobe Illustrator) – Adobe Illustrator proqramının istifadə etdiyi vektor tipli qrafik format;
- **BMP** (Bitmap) – MS Windows və ya OS/2 mühitində istifadə olunan rastr tipli qrafik format;
- **CDR** (Corel Draw) – Corel Draw proqramının istifadə etdiyi vektor tipli qrafik format;
- **GIF** (Graphics Interchange Format) – internetdə daha geniş yayılmış qrafik verilənlərin mübadiləsi üçün rastr tipli qrafik format;
- **JPEG/JPG/JPE** (Joint Photographic Experts Group) – əsasən rastr tipli təsvir (foto, şəkil və s.) üçün istifadə olunan və internetdə geniş yayılmış format;
- **MAX** – 3D Studio Max proqramının qrafik formatı;
- **PCX** – PaintBrush proqramı tərəfindən istifadə olunan rastr tipli qrafik format;
- **PNG** (Portable Network Graphics) – GIF faylların əvəzinə istifadə üçün nəzərdə tutulmuş fayl formatı;
- **PSD** – Photoshop təsvirlərin formatı;
- **TIFF/TIF** (Tagged Image File Format) – yüksək keyfiyyətli təsvir fayllarının teqləşdirilmiş rastr tipli qrafik formatı;
- **WDP/HDP** (Windows Media Photo) - Windows Media Photo proqramı tərəfindən istifadə olunan rastr tipli qrafik format;

- **WMF** (Windows Metafile) – Windows meta-faylları üçün nəzərdə tutulmuş format.

Səs faylları:

- **MIDI/MID** (Musical Instrument Digital Interface) – daxili sintezatorun idarə olunması üçün istifadə olunan proqramın formatı;
- **MP3** (MPEG Layer 3) – yüksək səviyyədə sıxılmış musiqi fayllarının saxlanması üçün istifadə olunan format;
- **RAM** (Real Audio Movie) – audio və video məlumatların axınla ötürülməsi üçün yaradılmış real audio-film formatı;
- **WAV** (WAVE sound file) – Microsoft şirkətinin musiqi faylları formatı;
- **WMA** (Windows Media Audio) – audioinformasiyanın saxlanması və translyasiyası üçün Microsoft şirkəti tərəfindən təklif olunmuş format.

Sənəd və mətn faylları:

- **DOC/DOCX/DOCM** – Microsoft Word proqramında yaradılmış sənəd formatları;
- **RTF** (Rich Text Format) – nişanlar qoyulmuş mətn sənədlərinin saxlanması üçün Microsoft şirkəti tərəfindən təklif olunmuş format;
- **TXT** – standart (ASCII) mətn sənədi formatı;
- **XLS/XLSX/XLSM** – Microsoft Excel proqramında yaradılmış elektron cədvəl formatı;
- **PPT/PPTX** – Microsoft Power Point proqramı tərəfindən yaradılan təqdimatların formatı.

Sənədlərin təsviri formatları:

- **EPS** (Encapsulated PostScript) – Kapsullaşdırılmış postskript formatı;

- **DVI.tex** (DeVice Independent) – kompüter tipografiyasında səhifələmə üçün TeX nəşriyyat sistemi tərəfindən istifadə olunan format;
- **PS** (PostScript) – postskript formatı.

Elektron kitab formatları:

- **DJVU** – kitabların skan surətini saxlanması üçün istifadə olunan format;
- **PDF** (Portable Document Format) – Adobe Acrobat proqramında yaradılmış sənəd, nəşr üçün kitabların səhifələnməsi üçün istifadə olunan format.

Hipermətn faylları:

- **HTML/HTM** (HyperText Markup Language) – HTML dilində internet səhifələrin yaradılması üçün istifadə olunan hipermətn formatı;
- **XML/XHTML** (eXtensible Markup Language) – XML dilində yaradılmış sənədlərin formatı.
- **MHTML/MHT** (MIME HyperText Markup Language) – HTML dilində arxivləşdirilmiş internet səhifələrin yaradılması üçün istifadə olunan hipermətn formatı;
- **CGI** (Common Gateway Interface) – internetdə proqram əlavələri üçün istifadə olunan format.

Proqram faylları:

- **EXE** – MS DOS və Windows mühitində yerinə yetirilən proqram faylları;
- **COM** – MS DOS və Windows mühitində sistem proqramları;
- **ASM** – Assembler dilində proqramın ilkin kodu;
- **C** – “C” dilində proqramın ilkin kodu;
- **CPP/CXX** – “C++” dilində proqramın ilkin kodu;
- **PAS** – Pascal dilində proqramın ilkin kodu;
- **BAS** – Basic dilində proqramın ilkin kodu.

III FƏSİL

İNFORMASIYA TƏLƏBATI

**İnformasiya tələbatı və onun məzmununa
təsir edən amillər**

İnformasiya tələbatının subyektləri

**İnformasiya tələbatının öyrənilməsi və
təqdim edilməsi üsulları**

İnformasiya tələbatının modeli

İnformasiya tələbatının formalaşması mərhələləri

İnformasiya tələbatının təsnifatı

**İnformasiya tələbatının tematik profilinin
müəyyən edilməsi**

İnformasiya ehtiyatının profilinin müəyyən edilməsi

**Fərdin bir mövzu sahəsi üzrə informasiya
tələbatının müəyyən edilməsi**

**Fərdin fəaliyyətinə aid olan bütün mövzu sahələri
üzrə informasiya tələbatının müəyyən edilməsi**

**Bir mövzu sahəsi üzrə fəaliyyət göstərən bütün
fərdlərin informasiya tələbatının məzmununun
müəyyən edilməsi**

Kollektiv informasiya tələbatının müəyyən edilməsi

3.1. İnformasiya tələbatı və onun məzmununa təsir edən amillər

Məlum olduğu kimi, hər bir peşəkar və ya qeyri-peşəkar fəaliyyət zamanı insanlarda müəyyən informasiyaya tələbat yaranır. İnformasiya tələbatının məzmunu fəaliyyət sahəsindən, insanların şəxsi, peşəkar və işgüzar maraqlarından və potensiallarından asılı olur. Belə ki, bir və ya bir necə sahəni əhatə edən peşəkar fəaliyyətlə məşğul olmaqla yanaşı, insanlar təhsil, ictimai-siyasi və idman fəaliyyəti, məişət işləri, hobbii və s. sahələrlə maraqlana bilər. Ona görə də onların ümumi informasiya tələbatı xarakterinə və məzmununa görə bir-birindən fərqlənən çoxlu sayda komponentlərdən ibarət olur.

Bir çox hallarda informasiyaya olan tələbat ailə, ətraf mühit, məktəb, ali təhsil müəssisələri, təşkilatlar və s. tərəfindən bu və ya digər dərəcədə təmin edilir, lakin bu biliklər real həyatda və praktiki fəaliyyətdə heç də həmişə kifayət etmir. Bəzən konkret məsələlərin həlli, o cümlədən problemlərin dərk olunması və aradan qaldırılması, qərarların qəbul edilməsi, yeni biliklərin alınması üçün mövcud məlumatlarla yanaşı əlavə, yeni məlumatlara ehtiyac yaranır.

Müxtəlif fəaliyyət sahələrində informasiya təminatı sistemlərinin fəaliyyətinin keyfiyyətini yaxşılaşdırmaq məqsədilə həyata keçirilən tədbirlər kompleksində istifadəçilərin peşəkar maraqlarının təmin olunmasına yönəlmiş informasiya ehtiyatlarının və xidmətlərinin spektrinin inkişaf etdirilməsi əhəmiyyətli yer tutmalıdır. İlk baxışda elə görünə bilər ki, müasir dövrdə informasiya təminatını reallaşdırılan ehtiyat və xidmətlərin spektri kifayət qədər

genişdir və mövcud informasiya tələbatının və ya müxtəlif istifadəçilər qruplarının peşəkar maraqlarının kifayət qədər əhatəli və dəqiq təmin edilməsinə imkan verir. Lakin təcrübə göstərir ki, istifadəçilərin informasiya tələbatlarının təmin edilməsi problemi indiyə qədər həll edilməmiş qalır.

Müasir dövrdə insanların (istifadəçilərin və ya abonentlərin) zəruri informasiya ilə keyfiyyətli və səmərəli şəkildə təmin edilməsində informasiya xidmətləri, o cümlədən kitabxana-informasiya sistemləri tərəfindən onların informasiya tələbatlarının öyrənilməsi böyük əhəmiyyət kəsb edir.

Ona görə də informasiya xidmətlərinin təşkili zamanı aşağıdakı amillər nəzərə alınmalıdır:

- Psixososial və psixoloji fenomen olan informasiya tələbatını təsvir etmək, xüsusilə də ölçmək çox çətinidir. Çox vaxt istifadəçilər öz tələbatlarını formalaşdırmaqda çətinlik çəkir və ya kifayət qədər dəqiq müəyyən edə bilmirlər. Axtarış sorğularında yol verilən leksik və ya semantik qeyri-müəyyənliklər tapılan məlumatların (axtarışının nəticəsinin) dəqiqliyi və dolğunluğuna əhəmiyyətli dərəcədə təsir edir.
- Müasir dövrdə informasiya tələbatı çoxaspektliyi və tez-tez dəyişməsi ilə xarakterizə olunur. Məsələn, alimlərin və mütəxəssislərin informasiya tələbatları mövzu baxımından və həcmcə kifayət qədər müxtəlif olur, elmi-texniki nailiyyətlərdən asılı olaraq dəyişir.
- İnformasiya tələbatları haqqında dəqiq məlumatın alınmasında çətinlik onların proqnoz xarakterli olması ilə əlaqədar olur. Çox vaxt problem istifadəçinin məlumatlılıq səviyyəsinin aşağı olması səbəbin-

dən obyektiv qeyri-müəyyənlik şəraitində informasiya tələbatının formalaşdırılması üçün zəruri məlumatların hazırlanması ilə bağlı olur.

- Axtarış sistemlərində informasiya-axtarış dillərinin xüsusiyyətləri informasiya tələbatının sorğular şəklində ifadə edilməsinin keyfiyyətinin, bunun nəticəsi olaraq axtarışın dəqiqliyinin əhəmiyyətli dərəcədə aşağı düşməsinə səbəb olur. Belə ki, sorğuların formalaşdırılması zamanı leksik (terminlərin birqiymətli olmaması), semantik (mövzunun məzmununun informasiya-axtarış dilində ifadəsinin qeyri-dəqiqliyi) və meta-informasiya qeyri-müəyyənliklərinin aradan qaldırılması zamanı obyektiv çətinliklər meydana çıxır.
- Müasir dövrdə informasiya ehtiyatlarının həcminin fasiləsiz artması, informasiyanın istehsalı və istehlakı sahəsində geniş imkanların yaranması bu sahədə mövcud problemlərə müasir ekoloji-informasiya vəziyyəti ilə bağlı olan kifayət qədər mürəkkəb problemlər əlavə edir.

Qeyd edilməlidir ki, informasiya tələbatları bir çox aspektlərdən öyrənilə və ya tədqiq oluna bilər:

- informasiya tələbatlarının məzmunu;
- informasiya tələbatlarının dəyişmə dinamikası;
- informasiya tələbatlarının meydanagəlmə forması və səbəbi;
- informasiya tələbatlarının meydanagəlmə intensivliyi və s.

İS-in fondlarının formalaşdırılmasının əsas amillərindən biri onların cəmiyyətin informasiya tələbatlarına uyğun olmasıdır. Başqa sözlə, İS-in fondlarının həcminin, tərkibinin və məzmununun optimal formalaşdırılması və

qiymətləndirilməsi üçün əsas meyar qismində cəmiyyətin, onun üzvlərinin informasiya tələbatları çıxış edir. Təcrübə göstərir ki, bir qayda olaraq, İS-in fondu ilə onun istifadəçilərinin informasiya tələbatları arasında müəyyən qarşılıqlı əlaqə mövcud olur. Belə ki, bir tərəfdən, İS-in fondunun keyfiyyətə dəyişməsi (fondun genişləndirilməsi, fonda yeni informasiya ehtiyatlarının daxil edilməsi, eləcə də köhnəlmiş, praktiki və nəzəri əhəmiyyətini itirmiş materialların fonda çıxarılması) onun istifadəçilərində mahiyyətə yeni informasiya tələbatlarının və sorğularının formalaşmasına gətirib çıxarır.

Digər tərəfdən, fəaliyyət sahələrinin genişlənməsi, informasiya potensialının artması, elm və texnikanın inkişafı və s. səbəblərdən istifadəçilərdə yeni informasiya tələbatlarının və sorğularının yaranması İS-in fonduna təsir göstərir, onun təkmilləşdirilməsi və genişləndirilməsi dəyişməsi zərurətini yaradır, inkişafını təmin edir.

Aydındır ki, İS-in əsas məqsədi onun istifadəçilərinin informasiya tələbatlarının tam ödənilməsindən ibarətdir. Bu məqsədin reallaşdırılması üçün müvafiq fond, optimal struktur və fəaliyyət mexanizmi tələb olunur. Bu baxımdan İS ilə onun istifadəçisi arasındakı qarşılıqlı əlaqədə başlıca rolunu onun informasiya tələbatı, marağı və sorğusu oynayır. Ona görə də “informasiya tələbatı”, “informasiya marağı” və “informasiya sorğusu” anlayışlarının qarşılıqlı təhlili və müqayisəsi maraqlı kəsb edir.

Qeyd edilməlidir ki, istifadəçinin informasiya sorğusuna münasibətdə ilkin olan informasiya tələbatı müəyyən mərhələdə konkret informasiya ehtiyatına yönəlmiş olur və sonradan həmin ehtiyata tələbat kimi qiymətləndirilə biləcək sorğuya çevrilir.

İnformasiya tələbatı – peşəkar fəaliyyətində və ya sosial-məişət təcrübəsində insanın qarşısında duran vəzifələrin yerinə yetirilməsi və məqsədin əldə edilməsi üçün tələb olunan əlavə informasiyaya olan tələbatdır. Elmin, təhsilin, istehsalın, texnikanın və s. inkişafı və təkmilləşməsi nəticəsində hər hansı informasiyaya, bu informasiyanı özündə saxlayan sənədlərə olan tələbat aradan qalxa və ya yeni informasiyaya tələbat yarana bilər.

İnformasiya marağı dedikdə tələb olunan informasiyanın müəyyənləşdirilməsi, daxil olan informasiyaya seçici münasibət, informasiya tələbatını potensial ödəyə biləcək kanalların və mənbələrin axtarılması başa düşülür. Bu amillər insanın informasiya potensialı ilə sıx bağlı olub, təlim-tərbiyə, təhsil, peşəkar fəaliyyət, gündəlik həyat və məişət qayğıları nəticəsində formalaşır. Məhz informasiya marağı vasitəsilə insanın şüurunda informasiya tələbatının təzahür etməsi prosesi baş verir, nəticədə dərk edilmiş informasiya tələbatı formalaşır.

İnformasiya sorğusu – insana fəaliyyətini həyata keçirmək üçün zəruri olan informasiya haqqında təsəvvürün (dərk edilmiş informasiya tələbatının) ifadəsidir. Təsəvvür kateqoriyasının həmişə subyektiv hal olduğunu nəzərə alaraq demək olar ki, bütün informasiya sorğuları, eləcə də dərk edilmiş informasiya tələbatları həmişə subyektiv xarakter daşıyır. Başqa sözlə, informasiya sorğusu istifadəçi tərəfindən özünün informasiya tələbatını ifadə etmək üçün tərtib etdiyi sözlər toplusudur. İnformasiya sorğusuna zəruri məlumatların əldə olunması üçün müvafiq qaydada tərtib edilərək İS-ə təqdim edilən giriş məlumatı kimi baxmaq olar.

İstifadəçi sorğusunu formalaşdırarkən yalnız özünün informasiya tələbatından çıxış edir. Lakin zəruri olan

informasiya haqqında kifayət qədər bilik olmadan sorğu yarana bilməz. İstifadəçinin lazım olan informasiya haqqında təsəvvürləri onun həmin anda müvafiq mövzu sahəsi üzrə biliyindən, malik olduğu informasiya potensialından bilavasitə asılı olur. Ona görə də informasiya sorğusunu bəzən insanın fəaliyyətində meydana çıxmış hər hansı məsələnin həlli üçün lazım olan, lakin çatışmayan zəruri informasiyanın “məlum olması haqqında bilik” adlandırırlar.

İS-in fondunun təkmilləşdirilməsinə təsir edən amillərdən biri də onun tərkibi və strukturu ilə bağlıdır. Aydındır ki, müxtəlif mövzu sahələrində insanlara forma və məzmun baxımından müxtəlif məlumatlar lazım olur. Lakin, bir qayda olaraq, peşəkar tələbatlar oxşar struktura malik olur. İnformasiya tələbatlarının strukturunun təhlili nəticəsində onun aşağıdakı əsas növləri fərqləndirilir.

Cari və retrospektiv informasiyaya tələbat. Peşəkar fəaliyyət sahəsində operativ qərarlar qəbul etmək, ixtisası üzrə və yaxın bilik sahələrində son nailiyyətlər səviyyəsindən geri qalmamaq üçün mütəxəssisdə daima yeni informasiyaya, yəni operativ cari məlumatlanmaya ehtiyac olur. Retrospektiv informasiyaya tələbat mütəxəssisin ixtisası və yerinə yetirdiyi vəzifələr ilə əlaqədar yaranır. Ona görə də bu növ tələbatlar nisbətən sabit olur və daim qüvvədə qalır. Bir çox hallarda təhlil aparmaq, verilmiş məsələ ilə bağlı əsaslandırılmış fikir formalaşdırmaq üçün keçmiş dövrə aid olan (retrospektiv) informasiyaya ehtiyac yaranır. Bu halda əvvəlki illərdə toplanmış mənbələrə müraciət olunur. Retrospektiv informasiyaya sorğular epizodik xarakter daşıyır və onları birdəfəlik sorğular adlandırırlar.

Dar və geniş mövzulu informasiyaya tələbat. İnformasiya fəaliyyəti prosesi daima tematik çərçivələrin

daralmasına gətirib çıxarır. Bunun da nəticəsində mütəxəssis tərəfindən konkret mövzu sahəsində informasiyaya tələbat yrana bilər. Belə informasiyanın istifadəsi ilk baxışda əhəmiyyətsiz görünən bir çox amillərin nəzərdən kənar qalmasına səbəb ola bilər. Ona görə də konkret mövzu sahəsi ilə tanış olmaq məqsədilə nisbətən geniş mövzulu informasiyanın təqdim edilməsi tövsiyə olunur.

Sahəvi (ixtisaslaşmış) və sahələrarası informasiyaya tələbat. Əksər aktual problemlər konkret məsələ üzrə sahəvi və sahələrarası informasiya ilə tanışlıq və onun təhlili yolu ilə həll olunur. Ona görə də belə informasiyaya tələbat daim aktual olur.

Faktoqrafik və konseptual informasiyaya tələbat. Bəzən istehlakçıda digər mütəxəssislərin və istehlakçıların məlumatları və qiymətləndirmələri barədə məlumata tələbat yaranır. Bunun üçün konseptual (mütəxəssis və istifadəçilərin rəyləri) və ya faktoqrafik (verilmiş sorğu üzrə konkret faktları və epizodları) informasiyadan istifadə edirlər.

3.2. İnformasiya tələbatının subyektləri

Peşəkar fəaliyyətində bu və ya digər dərəcədə informasiya tələbatı yaranacaq insanların əsas qruplarını qaçaqcadan müəyyənləşdirmədən informasiya xidmətinə tələb olunan səviyyədə təşkil etmək mümkün deyil. İnformasiya istehlakçılarının təsnif edilməsi üçün müxtəlif yanaşmalardan istifadə olunur. İlk öncə informasiya xidmətinin istifadəçilərini informasiyanın real və potensial istehlakçıları kimi kateqoriyalara bölmək olar.

Şübhəsiz ki, informasiya tələbatlarına insanların sosial-demokratik, şəxsi və işgüzar keyfiyyətləri (təhsil səviyyəsi, iş stajı, yaşı, həyat təcrübəsi və s.) təsir edir. Ona görə də informasiya xidmətinin təşkili zamanı istifadəçilərin peşəkar fəaliyyət sahələrini, ixtisaslarını, vəzifə statuslarını, funksional vəzifələrini və peşəkarlıq səviyyələrini nəzərə almaq vacibdir.

Ümumiyyətlə, peşəkar fəaliyyətdə bu və ya digər dərəcədə informasiya tələbatı yarana biləcək subyektlərə (informasiya istehlakçılara) müəyyən elmi problemlər üzərində tədqiqatlar aparan alimləri, istehsal sahələrində çalışan mütəxəssisləri, təhsil müəssisələrinin elmi-pedaqoji heyətini, orada təhsil alanları (tələbələr, şagirdləri və s.), idarə rəhbərlərini, menecerləri, layihə iştirakçılarını, rəhbərlərini, əlaqələndiricilərini, sahibkarları, kommersiya işçilərini, qulluqçuları, fəhlələri və s. aid etmək olar.

Praktikada informasiya xidmətinin həyata keçirilməsi zamanı subyektləri (informasiya istehlakçılarını) əsas bir neçə qrupa bölürlər. Birinci qrupa səviyyəsindən asılı olmayaraq bütün *rəhbərlər* daxil edilir. Daha yüksək ixtisaslı mütəxəssislər olan rəhbərlər vaxt azlığı şəraitində qərar qəbul etmək zərurəti ilə üzləşirlər. Təbii ki, qəbul edilmiş qərarların optimallığı rəhbərlərin məlumatlılığından çox asılı olur. Rəhbərin məlumatlılıq dərəcəsi nə qədər yüksək olarsa, qəbul edilən qərarların təsir dairəsi bir o qədər miqyaslı, onların informasiya tələbatı isə çox aspektli olar. Rəhbər işçilərin informasiya tələbatlarının tematik diapazonu genişliyi ilə fərqlənir.

Informasiya istehlakçılarının ikinci qrupunu *mütəxəssislər* təşkil edir. Bu kateqoriyaya müxtəlif fəaliyyət sahələrində, o cümlədən elm, təhsil, mədəniyyət və s. sahələrində çalışan, öz peşəkar fəaliyyətində xüsusi, o cümlədən

elmi-texniki və digər informasiyaya ehtiyacı yaranan bütün şəxslər aid edilir. Mütəxəssislərin ixtisası və peşəsi onların informasiya tələbatlarının tematik məzmununu müəyyən edir. Belə ki, onların informasiya tələbatlarının məzmununa, həmçinin, fəaliyyət göstərdikləri elm, təhsil və ya istehsalat sahələri, iş yeri, vəzifələri və s. əhəmiyyətli dərəcədə təsir göstərir. İxtisas çərçivəsində mütəxəssisləri onların yerinə yetirdikləri işlərin xarakterinə, tutduqları vəzifələrə və icra etdikləri funksiyalara görə təsnif etmək mümkündür. Məsələn, mütəxəssisləri elmi-tədqiqat işçilərinə – alimlərə, pedaqoji işçilərə – müəllimlərə, istehsalatda çalışan təcrübi mütəxəssislərə, idarəetmə fəaliyyəti ilə məşğul olan inzibatçılara və s. ayırmaq olar.

Informasiya istehlakçılarının növbəti qrupuna bilavasitə istehsalat və xidmət sahələrində işləyən *mühəndis-texniki işçilər* daxildir. Onlar informasiya istehlakçılarının daha böyük hissəsini təşkil edir. Bu kateqoriyadan olan mütəxəssislərin informasiya tələbatlarının xarakterik xüsusiyyəti nisbi stabilliyi ilə müəyyən olunur. Lakin elmi-texniki tərəqqinin təsiri altında bir çox fəaliyyət sahələrində, ələlxüsus sənayedə texnoloji proseslərin yeniləşməsi, köhnəlmiş avadanlıqların, maşınların, qurğuların, ləvazimatların, ilkin xammalın yeniləri ilə əvəzlənməsi, yeni informasiya texnologiyalarının, o cümlədən elektron texnikanın tətbiqi informasiya tələbatlarının məzmununa əhəmiyyətli təsir göstərir və dəyişikliklər olunmasını tələb edir. Bəzi fəaliyyət sahələrində çalışan mütəxəssislər (həkimlər, iqtisadçılar, müəllimlər, aqronomlar, zootexniklər və s.) də yerinə yetirdikləri işlərin xarakterinə görə mühəndis-texniki işçilər qrupuna aid edirlər.

Informasiya istehlakçılarının sonuncu qrupunu *sahibkarlar, iş adamları, kommersiya işçiləri* təşkil edirlər. Onlarda

daha çox faktoqrafik informasiyaya tələbat yaranır. Lakin bu kateqoriyaya daxil olan şəxslərin informasiya marağı diapazonuna görə əhəmiyyətli dərəcədə fərqlənə bilər.

3.3. İnformasiya tələbatının öyrənilməsi və təqdim edilməsi üsulları

İnformasiya tələbatının öyrənilməsi üsullarını ümumi halda iki qrupa bölmək olar:

- birbaşa üsullar;
- əlavə (köməkçi) üsullar

Birbaşa üsullar – subyektin qarşısında duran vəzifələrin, onun həll etdiyi məsələlərin sistemli təhlilinə əsaslanır və obyektiv zəruri olan informasiyanı aşkar etmək üçün onun fəaliyyət göstərdiyi konkret vəziyyətin öyrənilməsinə yönəlmiş olur.

Əlavə (köməkçi) üsullar – anketləşdirmə, istifadəçi istinadlarının təhlili, müsahibə götürmək və s. yolu ilə mövzu sahəsində istifadəçilərin fikirlərinin öyrənilməsinə əsaslanır. Bu üsulların çatışmazlığı ondan ibarətdir ki, informasiya tələbatları haqqında mülahizə insanların subyektiv fikirləri və ya onların məhdud praktiki təcrübəsi əsasında qurula bilər.

İnformasiya tələbatının öyrənilməsinin funksional vəzifələrin təhlilinə əsaslanan üsulları subyektin xidməti vəziyyətinə uyğun və peşəkar fəaliyyət sahəsində qərarların qəbul edilməsi üçün zəruri olan informasiya ilə təmin olunmasından ibarətdir.

Subyektin funksional vəzifələrinin sistemli təhlili əsasında onun fəaliyyəti üçün zəruri olan biliklərin həcmi müəyyən edilir. Burada sistemli təhlilin aparılması çox

vacibdir. Bir çox hallarda fərdin müəyyən sahədə fəaliyyəti digər sahələrlə bağlı olur. Bilavasitə vəzifələrini yerinə yetirmək üçün ona yalnız dar mövzuda (bir mövzu sahəsində) deyil, eyni zamanda qeyri-profil sahələrinə aid böyük həcmdə məlumatlar lazım olur.

Subyektin sahələrarası bu əlaqələri heç də həmişə düzgün təsəvvür etmirlər. Bir qayda olaraq, funksional vəzifələrin sistemli təhlili nəticəsində belə əlaqələr aşkar olunur. Bunların da əsasında subyektin qarşısında duran, həlli zəruri olan məsələlər müəyən edilir və subyektə lazım ola biləcək informasiya ehtiyatlarının (məlumatların, sənədlərin və s.) siyahısı formalaşdırılır. Nəticədə, subyektin fəaliyyətinin informasiya modeli (“informasiya pasportu”) qurulmuş olur.

Fəaliyyətin informasiya modeli özündə real informasiya tələbatını əks etdirir. Qeyd olunduğu kimi, subyekt lazım olan informasiyanı əldə etmək üçün zəruri sorğunu özünün dərk etdiyi informasiya tələbatı əsasında formalaşdırır və İS-ə təqdim edir. Əksər hallarda dərk edilmiş informasiya tələbatı real informasiya tələbatına uyğun olmur, ona görə də istifadəçinin sorğusu reallığı tam dolğun və düzgün əks etdirmir.

3.4. İnformasiya tələbatının modeli

Ümumiyyətlə, istənilən fəaliyyəti aşağıdakı dördlük vasitəsilə ifadə etmək olar:

$$F = \{S, P, A, V\}, \quad (3.1)$$

burada S – fəaliyyəti yerinə yetirən subyekt, P – fəaliyyətin yönəldiyi obyekt, yerinə yetirildiyi mövzu sahəsi, A – fəaliyyətin məqsədi, V – fəaliyyətin həyata keçirilməsi vasitələri, mexanizmləridir.

Burada subyekt dedikdə formasından, növündən, mövzu sahəsindən, əhatəliliyindən və təyinatından asılı olmayaraq hər hansı informasiyaya tələbatı yaradan fəaliyyəti həyata keçirən insanlar, fərdlər, kollektivlər (müəyyən fəaliyyət və ya maraq sahəsi üzrə insan qrupları), təşkilatlar və s. başa düşülür.

(3.1) modelinə daxil olan elementlər arasında əlaqələr fəaliyyətin subyektini vasitəsilə həyata keçirilir. Belə ki, subyekt fəaliyyətinin məqsədini müəyyən mexanizmlərdən istifadə etməklə yerinə yetirdiyi hərəkət və əməllər sistemi vasitəsilə obyektə (mövzu sahəsində) reallaşdırır.

Subyekt fəaliyyətini həyata keçirməzdən əvvəl, bir qayda olaraq, onu öz şüurunda planlaşdırır və onun informasiya modelini qurur. Bu model informasiya xarakterli olub, subyektin malik olduğu biliklər sistemi əsasında onun şüurunda formalaşır. İnformasiya modelində fəaliyyətin obyektiv şərtlərinin qabaqcadan əks olunması və subyektiv obrazının qurulması onun məqsədyönlü idarə olunmasını təmin edir. Ona görə də fəaliyyətin uğurla nəticələnməsi və qarşıya qoyulan məqsədin əldə olunması subyektin zəruri informasiyaya malik olub-olmaması dərəcəindən, yəni onun potensialından bilavasitə asılı olur. Nəticə etibarlı ilə insanın şüuru fəaliyyətin idarə edilməsinin başlıca mexanizminə çevrilir, vadaredici, nizamlayıcı və nəzarətedici kimi funksiyaları yerinə yetirir.

Əksər hallarda şüurda fəaliyyətin tam informasiya modelinin qurulması müəyyən çətinliklərlə qarşılaşır, müvafiq informasiya təminatı və əlavə istiqamətlənmə olmadan onun həyata keçirilməsi mümkün olmur. Ona görə də subyekt öz fəaliyyəti üçün çatışmayan informasiyanın əldə olunması istiqamətində tədbirlər görmək zərurətində qalır.

3.5. İnformasiya tələbatının formalaşması mərhələləri

Qeyd edildiyi kimi, informasiya tələbatı hər hansı fəaliyyət prosesində meydana çıxır və onun məzmununa bu fəaliyyəti həyata keçirən subyektlərin sosial-demoqrafik, şəxsi və işgüzar keyfiyyətləri (təhsil səviyyəsi, dünyagörüşü, iş və həyat təcrübəsi, şəxsi maraqları və s.) təsir edir.

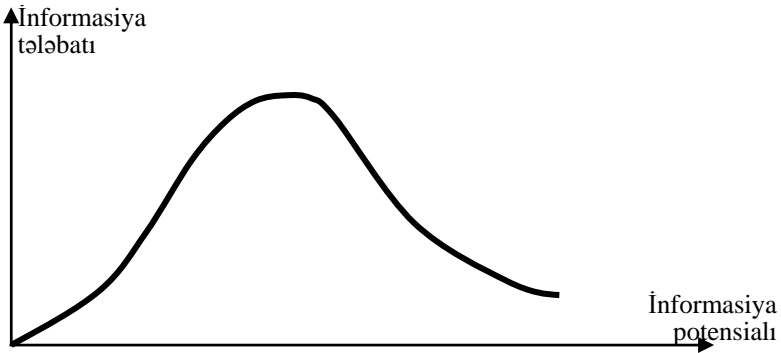
Qeyd etmək lazımdır ki, insanın istənilən fəaliyyəti onda informasiya tələbatının formalaşmasına gətirib çıxarmır. Bu, hər şeydən əvvəl fəaliyyətin xarakterindən, obyektindən, məqsədindən və subyektin informasiya potensialından asılı olur. Əgər subyekt qarşısında duran məqsədi reallaşdırmaq üçün əvvəllər həyata keçirdiyi fəaliyyəti təkrar edirsə və ya obyekt ona tam (və ya qismən) məlumdursa, onda əlavə informasiyanın alınmasına zərurət az olacaq və ya müəyyən müddət ərzində belə zərurət ümumiyyətlə aradan qalxacaqdır. Başqa sözlə, hər hansı peşəkar və ya qeyri-peşəkar fəaliyyət təkrarlanan xarakter daşıyarsa, əvvəllər əldə olunmuş informasiyaya və həyat təcrübəsinə əsaslanırsa, subyekt kifayət qədər biliyə və ya təcrübəyə malikdirsə, onda yeni informasiyanın alınmasına tələbat minimal olur.

Ümumiyyətlə, informasiya tələbatının formalaşmasında əsas üç amil mühüm rol oynayır:

- ***tələbatın mövcudluğu*** – fəaliyyətlə bağlı biliklərə ehtiyacın yaranması, tələb olunan biliklərin formalaşması üçün informasiya çatışmazlığının hiss olunması və onun aradan qaldırılmasına səy göstərilməsi;
- ***tələbatın məzmunu*** – tələb olunan informasiyanın tematikası, məzmunu, forması tutumu, eləcə də mövcudluğu barədə məlumatın olması;

- **informasiya potensialı** – informasiya çatışmazlığının subyekt tərəfindən dərk edilməsi, lazım olan (çatışmayan) informasiyanın müəyyənləşdirilməsi üçün zəruri biliklərin olması.

Qeyd edilməlidir ki, zəruri informasiya haqqında təsəvvür fəaliyyətin obyektı və şərtləri haqqında subyektə olan biliklər sistemi, yəni subyektin informasiya potensialı əsasında formalaşır. İnformasiya potensialı hər şeydən əvvəl obyektə bağlı olan informasiya tələbatının məzmununa təsir edir. Subyektin informasiya potensialı ilə onun informasiya yönümü arasında birbaşa asılılıq mövcuddur. Subyektin informasiya potensialı nə qədər geniş və dərindir, onun informasiya tələbatı bir o qədər dəqiq və konkret olur (şəkil 3.1).

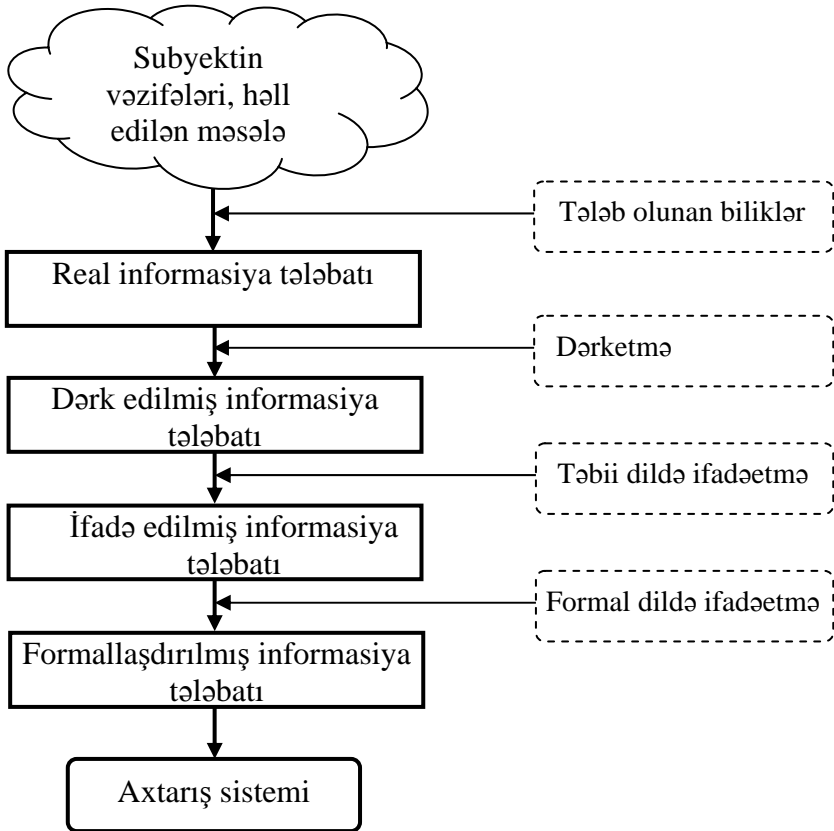


Şək.3.1. Subyektin informasiya tələbatının onun informasiya potensialından asılılığı

Müvafiq informasiya potensialının olmaması subyektə zəruri informasiya haqqında qeyri-müəyyən və qeyri-adekvat təsəvvürlərin formalaşmasına gətirib çıxara bilər ki, bunun da nəticəsində informasiya potensialları fərqli olan, lakin tamamilə eyni (məqsədi, obyektı və vasitələri eyni olan) fəaliyyətlə məşğul olan ayrı-ayrı subyektlərdə məzmunca müxtəlif informasiya tələbatları yarana bilər.

Ümumiyyətlə, informasiya tələbatının formalaşması prosesi dörd mərhələdən keçir (şəkil 3.2):

Birinci mərhələdə real informasiya tələbatı mövcud olur. Bu, məsələnin həlli və ya vəzifənin yerinə yetirilməsi üçün zəruri olan biliklər sistemini formalaşdırma biləcək informasiya tələbatıdır. Bu informasiya tələbatı subyekt tərəfindən dərk edilməyə də bilər.



Şək.3.2. İnformasiya tələbatının formalaşma mərhələləri

İkinci mərhələdə qarşıda duran vəzifə (məsələ) subyekt tərəfindən başa düşüldükdən sonra onun şüurunda bu vəzifə ilə bağlı informasiya tələbatı formalaşmış olur. Belə tələbata dərk edilmiş informasiya tələbatı deyilir.

Praktikada, bir qayda olaraq, hər hansı məsələnin həlli üçün tələb olunan real informasiya tələbatı subyektin şüurunda formalaşan dərk edilmiş informasiya tələbatından fərqlənir. Başqa sözlə, subyektin müvafiq sahədə məlumatlılıq dərəcəsi (potensialından) asılı olaraq, aşağıdakı vəziyyətlərdən biri yarana bilər:

- subyekt məsələnin həlli üçün hansı informasiyanın zəruri olduğunu bilmir;
- lazım olan informasiyanın bir hissəsi artıq subyektə vardır;
- subyekt bu sahədə müəyyən biliklərə malikdir (yüksekixtisaslı mütəxəssis və ya təcrübəli kadrdır).

Üçüncü mərhələdə subyekt tərəfindən dərk edilmiş informasiya tələbatının təbii dilin köməyi ilə təsvir olunması nəticəsində *ifadə edilmiş informasiya tələbatı* yaranır.

Dördüncü mərhələdə zəruri informasiyanın axtarılması məqsədilə ifadə edilmiş informasiya tələbatı informasiya xidməti, o cümlədən informasiya-axtarış sisteminin formal axtarış vasitələrinin (informasiya-axtarış dilinin) köməyi ilə təqdim edilir və nəticədə *formallaşdırılmış informasiya tələbatı* meydana gəlir. Formallaşdırılmış informasiya tələbatı əsasında axtarış sorğusu tərtib olunur.

Beləliklə, *informasiya tələbatı* dedikdə fəaliyyətin yerinə yetirilməsi prosesində çatışmayan biliklərin əldə olunması üçün subyektin zəruri olan informasiyaya dərk edilmiş tələbatı başa düşülür. Ümumi halda, bu tələbatı fəaliyyət sahəsi üzrə tələb olunan biliklərlə subyektin malik olduğu biliklər arasındakı fərq kimi ifadə etmək olar:

$$IT = TB - MB, \quad (3.2)$$

burada IT – subyektin dərk edilmiş informasiya tələbatı, TB – fəaliyyət sahəsi üzrə funksional vəzifələrinin yerinə yetirilməsi üçün tələb olunan biliklər, real informasiya tələbatı, MB – fəaliyyət sahəsi üzrə subyektin malik olduğu biliklərdir.

3.6. İnformasiya tələbatının təsnifatı

İnformasiya tələbatlarını müxtəlif meyarlara görə təsnif etmək olar, lakin daha çox aşağıdakı yanaşmalara rast gəlinir.

İnformasiya tələbatının əhatə dairəsinə görə:

- fərdi informasiya tələbatları – ayrı-ayrı insanlarda formalaşan informasiya tələbatlarıdır;
- kollektiv informasiya tələbatları – müəyyən maraq dairəsi (əmək fəaliyyəti, elm, mədəniyyət, hobbisi və s.) ətrafında birləşən insanlar qrupunun kollektiv şüurunda formalaşan informasiya tələbatıdır.

Kollektiv informasiya tələbatları kollektivi təşkil edən fərdlərin informasiya tələbatları əsasında formalaşır. Lakin kollektiv informasiya tələbatlarını fərdi informasiya tələbatlarının sadə toplanması kimi qəbul etmək olmaz. Kollektiv informasiya tələbatı həmin kollektivə daxil olan ayrı-ayrı fərdlərin informasiya tələbatlarının yalnız kollektivin ümumi məqsəd və vəzifələrinə uyğun olan hissələrinin toplanması nəticəsində formalaşır.

İnformasiya tələbatının reallaşma səviyyəsinə görə:

- real informasiya tələbatı – hər hansı məsələnin həlli və ya vəzifənin yerinə yetirilməsi üçün tələb olunan informasiya;

- dərk edilmiş informasiya tələbatı - hər hansı məsələ subyekt tərəfindən dərk edildikdən sonra onun həlli üçün tələb olunan informasiya.

İnformasiya tələbatının ifadə dərəcəsinə görə:

- passiv informasiya tələbatları;
- ifadə olunmuş informasiya tələbatları.

İnformasiya tələbatının mövcudolma müddətinə görə:

- müvəqqəti informasiya tələbatları;
- daimi informasiya tələbatları.

İnformasiya tələbatının formalaşma sahəsinə görə:

- peşəkar informasiya tələbatları;
- qeyri-peşəkar informasiya tələbatları.

Tələb olunan informasiyanın mövzusunə (tematikaya) görə:

- dar mövzulu informasiya;
- geniş mövzulu informasiya.

Tələb olunan informasiyanın növünə görə:

- faktoqrafik informasiya;
- konseptual informasiya;
- elmi informasiya;
- metodik informasiya;
- təlimat xarakterli informasiya və s.

Tələb olunan sənədlərin növünə görə:

- kitablar;
- jurnallar;
- xüsusi elmi-texniki sənədlər;
- hesabatlar və s.

İnformasiya istehlakçılmasına görə:

- rəhbərlər;
- alimlər;

- mühəndislər;
- sahibkarlar və s.

3.7. İnformasiya tələbatının tematik profilinin müəyyən edilməsi

Tutaq ki, $T = \{t_1, t_2, \dots, t_f\}$ – məlumatların (ehtiyatların), mövzu sahələrinin tematik profillərini müəyyən edən əlamətlər (terminlər, açar sözlər) çoxluğudur. Onda P_k fəaliyyət sahəsinin tematik profilini (mövzu sahəsini) T çoxluğunun elementləri vasitəsilə aşağıdakı kimi təsvir etmək olar:

$$P_k \Rightarrow T_k^P = \left\{ \left(t_q, \omega_k^P(t_q) \right) \mid t_q \in T, q = \overline{1, f} \right\},$$

$$k = \overline{1, m}. \quad (3.3)$$

Burada $\omega_k^P(t_q): T \rightarrow [0,1]$ – t_q termininin T_k^P çoxluğuna mənsubiyyət funksiyası olub, t_q termininin P_k mövzu sahəsinin profilinə aid olması dərəcəsini müəyyən edir. Mənsubiyyət funksiyasına qiymətlər mövzu sahələrinin profilləri müəyyən edilərkən qabaqcadan verilir.

Analoji qaydada, S_j subyektinin informasiya tələbatını da T çoxluğunun elementləri ilə aşağıdakı kimi təsvir etmək olar:

$$IT_j \Rightarrow T_j^{IT} = \left\{ \left(t_q, \eta_j(t_q) \right) \mid t_q \in T, q = \overline{1, f} \right\},$$

$$j = \overline{1, n}. \quad (3.4)$$

Burada $\eta_j(t_q): T \rightarrow [0,1]$ – t_q termininin S_j subyektinin informasiya tələbatının profilini təsvir edən T_j^{IT} çoxluğuna mənsubiyyət funksiyası olub, t_q termininin S_j subyektinin

informasiya tələbatının profili üçün vaciblik dərəcəsini göstərir. Onun qiyməti subyekt tərəfindən daxil edilə və ya ekspert qiymətləndirilməsi yolu ilə müəyyən edilə bilər. T_j^{IT} çoxluğu subyekt tərəfindən daxil edilmiş açar sözlərə, onun funksional vəzifəsinin və həll etdiyi məsələlərini mövzu sahələrinin, onun haqqında doldurulmuş anketlərdə göstərilən maraq dairəsinin təhlilinə əsasən formalaşdırılır.

Qeyd olunanlara əsasən, S_j subyektinin informasiya tələbatının məzmununu $P_k, k = \overline{1, m}$ mövzu sahələrinin profillərinə yaxınlıq dərəcələrini tapmaqla müəyyən etmək olar. Bunun üçün əvvəlcə, ayrı-ayrı $t_q, q = \overline{1, f}$ terminlərinə münasibətdə hər bir $IT_j, j = \overline{1, n}$ və $P_k, k = \overline{1, m}$ cütləri üçün onların relevantlıq dərəcələri hesablanır. Bu məqsədlə T_j^{IT} və T_k^P çoxluqlarının kəsişməsini tapmaq lazımdır:

$$T_j^{IT} \cap T_k^P, j = \overline{1, n}, k = \overline{1, m}. \quad (3.5)$$

Başqa sözlə,

$$\begin{aligned} \mu_{jk}(t_q) &= \eta_j(t_q) \times \omega_k^P(t_q), j = \overline{1, n}, k = \overline{1, m}, \\ q &= \overline{1, f}. \end{aligned} \quad (3.6)$$

Qeyd olunmalıdır ki, $\mu_{jk}(t_q)$ funksiyasının ayrı-ayrı terminlər üçün qiymətlərinə görə informasiya tələbatının məzmunu ilə mövzu sahəsinin profili arasındakı relevantlıq dərəcələri haqqında fikir söyləmək çətindir. Belə ki, informasiya tələbatları və mövzu sahələrinin profilləri üçün ayrı-ayrı terminlərin vaciblik dərəcələri (əhəmiyyəti) fərqli olur. Yalnız bütün terminlərin vaciblik dərəcələri və bu dərəcələr arasındakı fərqlər nəzərə alınmaqla informasiya tələbatlarının və mövzu sahələrinin profilləri

arasında ümumi orta relevantlıq dərəcəsinə müəyyənləşdirmək mümkündür.

Bu məqsədlə $P_k, k = \overline{1, m}$ mövzu sahəsi üçün $t_q, q = \overline{1, f}$ terminlərinin hər bir cütü arasında $\lambda_k(t_q, t_{q'})$ qeyri-səlis üstünlük münasibətləri $\lambda_k(t_q, t_{q'}): T_k^P \times T_k^P \rightarrow [0, 1]$ müəyyən edilir. Onun qiymətləri t_q termininin $t_{q'}$ terminindən nə dərəcədə üstün olduğunu göstərir və terminlərin mövzu sahələri üçün çəki əmsallarına əsasən aşağıdakı kimi hesablanır:

$$\begin{aligned} & \lambda_k(t_q, t_{q'}) = \\ & = \begin{cases} 1 - [\omega_k^P(t_{q'}) - \omega_k^P(t_q)], & \text{əgər } \omega_k^P(t_q) < \omega_k^P(t_{q'}) \\ 1, & \text{əks halda.} \end{cases} \quad (3.7) \end{aligned}$$

Qeyd olunanlar, eləcə də bütün terminlər və onlar arasındakı qeyri-səlis üstünlük münasibətləri nəzərə alınmaqla informasiya tələbatlarının məzmununun mövzu sahələrinin profillərinə relevantlıq dərəcələrini aşağıdakı kimi müəyyən etmək olar:

$$\begin{aligned} \mu_k(IT_j) &= \max_{t_q} \left\{ \max_{t_{q'}} \{ \min \{ \mu_{jk}(t_q), \mu_{jk}(t_{q'}), \lambda_k(t_q, t_{q'}) \} \} \right\}, \\ j &= \overline{1, n}, k = \overline{1, m}. \quad (3.8) \end{aligned}$$

Burada $t_q \neq t_{q'}$. $\mu_k(IT_j)$ funksiyasının alınan qiymətləri S_j subyektinin informasiya tələbatının P_k mövzu sahəsinin profilinə yaxınlıq dərəcəsinə göstərir.

Beləliklə, S_j subyektinin informasiya tələbatının profili hər hansı çox kiçik $\varepsilon > 0$ sərhəd qiyməti üçün $\mu_k(IT_j) > \varepsilon$ şərtini ödəyən bütün P_k mövzu sahələrinin profillərinə relevant qəbul oluna bilər.

3.8. İnformasiya ehtiyatının profilinin müəyyən edilməsi

(3.4) və (3.7) ifadələrinə uyğun olaraq r_i informasiya ehtiyatlarını T çoxluğunun elementləri ilə aşağıdakı kimi təsvir etmək olar:

$$r_i \Rightarrow T_i^r = \left\{ (t_q, \rho_i(t_q)) \mid t_q \in T, q = \overline{1, f} \right\}, i = \overline{1, l}. \quad (3.9)$$

Burada $\rho_i(t_q): T \rightarrow [0, 1] - t_q$ termininin r_i informasiya ehtiyatını təsvir edən T_i^m çoxluğuna mənsubiyyət funksiyası olub, t_q termininin r_i informasiya ehtiyatı üçün vaciblik dərəcəsini göstərir. Onun qiymətini avtomatik indeksləşdirmə üsullarının köməyi ilə hesablamaq olar. Bunun üçün, bir qayda olaraq, xüsusi proqram təminatı (robot, pauk və s. proqramları) işlənib hazırlanır.

Qeyd olunanlara əsasən, $r_i, i = \overline{1, l}$ məlumatlarının $P_k, k = \overline{1, m}$ mövzu sahələrinin profillərinə aid olması, yəni R_k çoxluğuna yaxınlıq (relevantlıq) dərəcəsini müəyyən etmək üçün aşağıdakı üsuldən istifadə etmək olar.

Əvvəlcə, ayrı-ayrı $t_q, q = \overline{1, f}$ terminlərinə münasibətdə hər bir $r_i, i = \overline{1, l}$ və $P_k, k = \overline{1, m}$ cütləri üçün onların profillərinin relevantlıq dərəcələri hesablanır. Bunun üçün uyğun olaraq T_k^P və T_i^m çoxluqlarının kəsişməsini tapır:

$$T_k^P \cap T_i^m, k = \overline{1, m}, i = \overline{1, n}.$$

Başqa sözlə,

$$\gamma_{ik}(t_q) = \omega_k^P(t_q) \times \rho_i(t_q), i = \overline{1, l}, k = \overline{1, m}, \\ q = \overline{1, f}. \quad (3.10)$$

Qeyd etmək lazımdır ki, əvvəlki bölməyə analoji olaraq, mövzu sahələri ilə informasiya ehtiyatları arasında

orta çəkili relevantlıq dərəcəsini müəyyən etmək üçün $\lambda_k(t_q, t_{q'})$ qeyri-səlis üstünlük münasibətlərini istifadə etmək olar.

Yuxarıda qeyd olunanları nəzərə alaraq, r_i informasiya ehtiyatının P_k mövzu sahəsinə relevantlıq dərəcəsini aşağıdakı düsturun köməyi ilə müəyyən etmək olar:

$$\gamma_k(r_i) = \max_{t_q} \left\{ \max_{t_{q'}} \{ \min \{ \gamma_{ik}(t_q), \gamma_{ik}(t_{q'}), \lambda_k(t_q, t_{q'}) \} \} \right\},$$

$$i = \overline{1, l}, k = \overline{1, m}, t_q \neq t_{q'}. \quad (3.11)$$

Əgər $\varepsilon \in [0, 1]$ olarsa, onda $\gamma_k(r_i) > \varepsilon$ şərtini ödəyən bütün P_k mövzu sahələri r_i informasiya ehtiyatına relevant kimi qəbul edilə bilər.

3.9. Fərdin bir mövzu sahəsi üzrə informasiya tələbatının müəyyən edilməsi

Beləliklə, qeyd olunanlar nəzərə alınaraq, s_j subyektinin p_k mövzu sahəsi üzrə informasiya tələbatını müəyyənləşdirmək üçün R_k və $\overline{R_{jk}^S}$ çoxluqlarının kəsişməsini tapmaq lazımdır:

$$R_{jk}^T = R_k \cap \overline{R_{jk}^S}, i = \overline{1, l}, j = \overline{1, n}, k = \overline{1, m}, \quad (3.12)$$

Burada $\overline{R_{jk}^S} - R_{jk}^S$ çoxluğunun tamamlayıcısıdır. Onun mənsubiyyət funksiyası aşağıdakı kimi hesablanır:

$$\overline{\varphi_{jk}(m_i)} = 1 - \varphi_{jk}(m_i). \quad (3.13)$$

Əgər

$$R_{jk}^T = \left\{ (r_i, \alpha_{jk}(r_i)) \mid r_i \in R, i = \overline{1, l} \right\}, k = \overline{1, m}, \\ j = \overline{1, n}, \quad (3.14)$$

olarsa, onda qeyri-səlis çoxluqların kəsişməsi qaydalarına uyğun olaraq

$$\alpha_{jk}(r_i) = \min_{i=1, l} \left\{ \mu_k(r_i), (1 - \varphi_{kj}(r_i)) \right\}, j = \overline{1, n}, \\ k = \overline{1, m} \quad (3.15)$$

yazmaq olar. Burada $\alpha_{jk}(r_i): R \rightarrow [0, 1] - R_{jk}^T$ çoxluğunun mənsubiyyət funksiyası olub, s_j subyektinin p_k mövzu sahəsinə aid olan r_i informasiya ehtiyatına tələbatını göstərir.

Hər hansı çox kiçik $\varepsilon > 0$ sərhəd qiyməti üçün $\alpha_{jk}(r_i) > \varepsilon$ şərtini ödəyən bütün m_i informasiya ehtiyatları çoxluğu s_j subyektinin p_k mövzu sahəsi üzrə tələbatına uyğun informasiya ehtiyatları kimi qəbul oluna və ona təqdim edilə bilər.

3.10. Fərdin fəaliyyətinə aid olan bütün mövzu sahələri üzrə informasiya tələbatının müəyyən edilməsi

s_j subyektinin məşğul olduğu bütün mövzu sahələri üzrə ümumi informasiya tələbatını müəyyənləşdirmək üçün onun ayrı-ayrı mövzu sahələri üzrə informasiya tələbatlarını özündə saxlayan $R_{jk}^T, j = \overline{1, n}$ çoxluqlarının birləşməsini tapmaq lazımdır:

$$R_j^{ST} = \cup_{k=1}^m R_{jk}^T, j = \overline{1, n}. \quad (3.16)$$

Əgər

$$R_j^{ST} = \left\{ \left(r_i, \alpha_j^S(r_i) \right) \mid r_i \in R, i = \overline{1, l} \right\}, j = \overline{1, n},$$

olarsa, onda R_j^{ST} çoxluğunun mənsubiyyət funksiyasının $\alpha_j^S(r_i): R \rightarrow [0, 1]$ qiymətlərini

$$\alpha_j^S(r_i) = \max_{k=\overline{1, m}} \alpha_{jk}(r_i), j = \overline{1, n},$$

yəni

$$\alpha_j^S(r_i) = \max_{k=\overline{1, m}} \min_{i=\overline{1, l}} \left\{ \mu_k(r_i), \left(1 - \varphi_{kj}(r_i) \right) \right\},$$

$$j = \overline{1, n} \quad (3.17)$$

kimi hesablamaq olar.

Əgər çox kiçik müsbət $\varepsilon \in [0, 1]$ sərhəd qiyməti verilmişdirsə, onda $\alpha_j^S(r_i) > \varepsilon$ şərtini ödəyən bütün r_i informasiya ehtiyatları çoxluğunu s_j subyektinin informasiya tələbatına uyğun informasiya ehtiyatları kimi ona təqdim etmək olar.

3.11. Bir mövzu sahəsi üzrə fəaliyyət göstərən bütün fərdlərin informasiya tələbatının məzmununun müəyyən edilməsi

Eyni p_k mövzu sahəsi üzrə ümumi informasiya tələbatını, başqa sözlə, bu sahədə məşğul olan bütün $S_j, j = \overline{1, n}$ subyektlərinin ümumi informasiya tələbatını müəyyən-ləşdirmək üçün ayrı-ayrı subyektlərin həmin mövzu sahəsi üzrə informasiya tələbatlarını özündə saxlayan $R_{jk}^T, j = \overline{1, n}$ çoxluqlarının birləşməsini tapmaq kifayətdir:

$$R_k^{PT} = \bigcup_{j=1}^n R_{jk}^T, k = \overline{1, m}. \quad (3.18)$$

Onda

$$R_k^{PT} = \{(r_i, \alpha_k^P(r_i)) | r_i \in R, i = \overline{1, l}\}, k = \overline{1, m} \quad (3.19)$$

çoxluğunun $\alpha_k^P(r_i): R \rightarrow [0, 1]$ mənsubiyyət funksiyasının qiymətlərini hesablamaq üçün

$$\alpha_k^P(r_i) = \max_{j=1, n} \alpha_{jk}(r_i), k = \overline{1, m}, \quad (3.20)$$

yəni

$$\alpha_k^P(r_i) = \max_{j=1, n} \min_{i=1, l} \{\mu_k(r_i), (1 - \varphi_{kj}(r_i))\},$$

$$k = \overline{1, m} \quad (3.21)$$

yazmaq olar.

Hər hansı çox kiçik müsbət $\varepsilon \in [0, 1]$ sərhəd qiyməti üçün $\alpha_k^P(r_i) > \varepsilon$ şərtini ödəyən bütün r_i məlumatları çoxluğu p_k mövzu sahəsi üzrə ümumi informasiya tələbatına uyğun məlumatlar çoxluğu olar. Bu tələbat informasiya xidmətinin, o cümlədən müvafiq mövzu sahəsi üzrə informasiya təminatının təşkili zamanı nəzərə alın bilər.

3.12. Kollektiv informasiya tələbatının müəyyən edilməsi

Təşkilatda bütün mövzu sahələri üzrə fəaliyyət göstərən subyektlərin hamısının birlikdə ümumi informasiya tələbatını (kollektiv informasiya tələbatını) $R^T = \{(r_i, \alpha(r_i)) | r_i \in R, i = \overline{1, l}\}$ müəyyənləşdirmək məqsədilə $j = \overline{1, n}$ üçün R_j^{ST} və ya $k = \overline{1, m}$ üçün R_k^{PT} çoxluqlarının birləşməsini tapmaq lazımdır:

$$R^T = \bigcup_{j=1}^n R_j^{ST}, \quad (3.22)$$

yəni

$$\alpha(r_i) = \max_{j=1, \overline{n}} \alpha_j^S(r_i) \quad (3.23)$$

və ya

$$R^T = \bigcup_{k=1}^m R_k^{PT}, \quad (3.24)$$

yəni

$$\alpha(r_i) = \max_{k=1, \overline{m}} \alpha_k^P(r_i). \quad (3.25)$$

Burada R^T çoxluğunun mənsubiyyət funksiyası $\alpha(r_i): R \rightarrow [0,1]$ kimi təyin olunur və onun hesablanması üçün yekun ifadə

$$\alpha(r_i) = \max_{j=1, \overline{n}} \max_{k=1, \overline{m}} \min_{i=1, \overline{l}} \left\{ \mu_k(r_i), (1 - \varphi_{kj}(r_i)) \right\} \quad (3.26)$$

olar.

Hər hansı çox kiçik müsbət $\varepsilon \in [0,1]$ sərhəd qiyməti üçün $\alpha(r_i) > \varepsilon$ şərtini ödəyən bütün r_i məlumatları çoxluğu təşkilatın ümumi informasiya tələbatına (kollektiv tələbat) uyğun məlumatlar çoxluğu olar. Bu, təşkilatın rəhbərliyi tərəfindən ümumi informasiyaya olan tələbat kimi qəbul edilir, ayrı-ayrılıqda subyektlərin fərdi, kollektiv və bütövlükdə təşkilatın fəaliyyətinin informasiya təminatı üçün əsas kimi götürülə bilər.

IV FƏSİL

İNFORMASIYA AXTARIŞININ NƏZƏRİ ƏSASLARI

İnformasiya axtarışının əsas anlayışları

İnformasiya-axtarış dili

**İnformasiya axtarışının nəticəsinin
qiymətləndirilməsinin əsas meyarları**

4.1. İnformasiya axtarışının əsas anlayışları

Klassik mənada informasiya axtarışı hər hansı statistik sənədlər kolleksiyasından axtarış sorğusuna uyğun olan sənədlərin tapılmasından ibarətdir. Bura modelləşdirmə, sənədlərin təsnif edilməsi, süzgəcdən keçirilməsi və klastertləşdirilməsi, axtarış sistemlərinin arxitekturasının və istifadəçi interfeyslərinin layihələndirilməsi, sənədlərin xülasələrinin və referatlarının hazırlanması, informasiyanın çıxarılması, informasiya-axtarış dilləri və s. məsələlər aid edilir.

Geniş mənada informasiya axtarışı aşağıdakı əsas vəzifələrin həllinin təmin edilməsinə yönəlmiş olur:

- Sorğuya relevant, informasiya tələbatını ödəyən informasiyanın axtarılıb tapılması;
- bir-birinə yaxın sahələrdə analoji informasiyanın axtarılması (ilkin informasiyanın genişlənməsi);
- əldə olunmuş informasiyanın ümumiləşdirilməsi və dəqiqləşdirilməsi;
- bilavasitə real məsələlərdən çıxış edərək informasiyanın təhlili və qiymətləndirilməsi.

İnformasiya axtarışı probleminin mürəkkəbliyi informasiyanın mənbələr çoxluğu üzrə paylanması qanununun obyektiv xarakteri ilə izah olunur. Belə ki, bu problem informasiya infrastrukturunun mürəkkəbliyi, hər hansı fəaliyyət sahəsi üzrə insanlara, o cümlədən mütəxəssislərə lazım olan bütün sənədli informasiya mənbələrinin bir yerə yığılmasının qeyri-mümkünlüyü, eləcə də müasir dövrdə mövcud olan informasiya axınlarının yuxarıda qeyd olunan xüsusiyyətləri ilə bağlı olur.

İnformasiya axtarışı problemi bu və ya digər dərəcədə informasiya-kommunikasiya texnologiyalarının, o cümlədən internetin informasiya ehtiyatlarının köməyi ilə həll edilir. Lakin informasiya çatışmazlığı problemini internet tamamilə həll edə bilməz. Çünki internetdə tapılmış informasiyanın dolğunluğu, doğruluğu və həqiqiliyi meyarlarının qiymətləri heç də həmişə müsbət olmur.

İnformasiya axtarışı (ingiliscə “information retrieval”) termini ilk dəfə 1952-ci ildə Kelvin Muyers tərəfindən istifadə olunmuşdur.

Ümumi halda, *informasiya axtarışı* (İA) dedikdə hər hansı sənədlər çoxluğundan göstərilən mövzuya aid olan, əvvəlcədən müəyyən edilmiş axtarış şərtini (axtarış sorğusunu) ödəyən zəruri faktları, məlumatları və verilənləri özündə saxlayan sənədlərin tapılması prosesi başa düşülür. Başqa sözlə, İA informasiya tələbatının ödənilməsi üçün istifadəçinin təqdim etdiyi axtarış sorğusuna uyğun olan məlumatların (və ya bu məlumatları özündə saxlayan sənədlərin) tapılması prosedurasıdır.

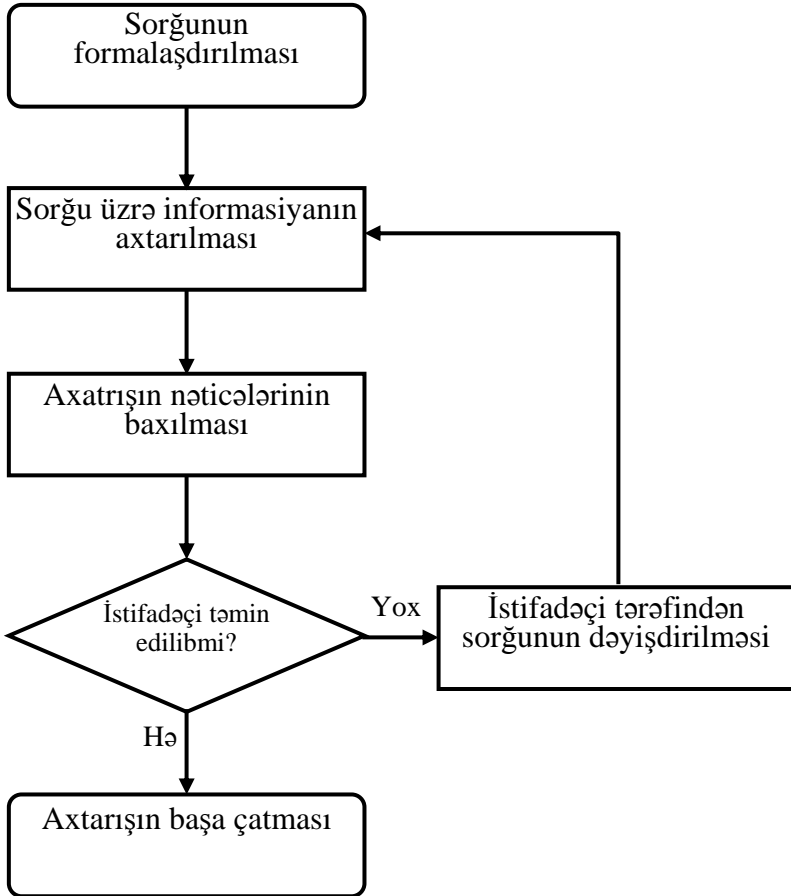
İA prosesi informasiyanın toplanması, emalı, aidiyyəti şəxslərə təqdim edilməsinə yönəlmiş əməliyyatlar ardıcılığını özündə birləşdirir. İA zəif strukturlaşdırılmış verilənlərin təqdim edilməsi, saxlanması və təşkili məsələləri ilə bilavasitə bağlı olan prosesdir. İA-nın elementar vahidi sənəddir.

İS prosesi dörd mərhələdə yerinə yetirilir:

- informasiya tələbatının müəyyən edilməsi və informasiya sorğusunun formalaşdırılması;
- mümkün informasiya mənbələri (massivləri) məcmusunun müəyyən edilməsi;
- aşkar olunmuş informasiya massivlərindən informasiyanın seçilib götürülməsi və təqdim edilməsi;

- alınmış informasiya ilə tanışlıq və axtarışın nəticəsinin qiymətləndirilməsi.

İA prosedurası ümumi şəkildə sxematik olaraq 4.1 sayılı şəkildə göstərilmişdir.



Şək.4.1. İnformasiya axtarışı prosedurası

Şəkildən görüldüyü kimi, informasiya axtarışı prosesi iterativ proseduradır, yəni istifadəçinin tələbləri ödənilmədikdə ona sorğusunda dəyişikliklər etmək və yeni sorğu üzrə təkrar axtarış aparmaq imkanı verilir. İstifadəçinin tələbatı tam ödənilənə qədər bu proses davam etdirilə bilər.

İnformasiya axtarışının əsas məqsədi istifadəçilərə onların informasiya tələbatlarının ödənilməsində yardım etməkdən ibarətdir. İstifadəçinin informasiya tələbatı axtarılan informasiyanı xarakterizə edən açar sözlər çoxluğundan ibarət olan sorğu şəklində formalaşdırılır. Lakin məlumdur ki, informasiya tələbatını dolğun şəkildə əks etdirən axtarış sorğusunun təsvir edilməsi texniki baxımdan o qədər də asan məsələ deyildir.

Ümumiyyətlə, informasiya axtarışının həyata keçirilməsinin əsas məqsədi aşağıdakılardan biri ola bilər:

- informasiya mənbəyi haqqında zəruri məlumatların axtarılması və onun digər mənbələr sistemində mövcud olmasının müəyyən edilməsi – daha effektiv axtarışın aparılması üçün yaradılmış bibliografik informasiyanın və bibliografik vəsaitlərin istifadəsi;
- lazım olan informasiyanı özündə saxlayan informasiya mənbələrinin (sənədlərin və nəşrlərin) özlərinin axtarılması;
- nəşrlərdə olan faktiki (məsələn, tarixi faktlar və hadisələr, konkret obyektlərin arxitektura qiymətliliyi, maşın, qurğu və proseslərin texniki xarakteristikaları, maddə və materialların xassələri, alimin, mühəndisin bibliografik verilənləri haqqında) məlumatların axtarılması.
- müxtəlif miqyaslı, ətraflı, dərin və s. tədqiqatlara əsaslanan ümumi xarakterli analitik informasiyanın axtarılması.

Bu məqsədlərə uyğun olaraq, informasiya axtarışının dörd əsas növünü fərqləndirirlər:

- bibliografik axtarış;
- sənədli axtarış;
- faktoqrafik axtarış;
- analitik axtarış.

Eyni zamanda axtarış obyektinin növündən, formasından, məzmunundan və formatından asılı olaraq, informasiya axtarışının aşağıdakı növləri qeyd olunur:

- ***Tammətnli axtarış*** – sənədlərin məzmunu (bütün mətni) üzrə axtarış prosesidir. Tammətnli axtarış sistemlərinə nümunə kimi internetin informasiya-axtarış sistemlərini göstərmək olar.
- ***Meta-verilənlər üzrə axtarış*** – sənədlərin axtarış sistemi tərəfindən qəbul edilən atributlarına və ya rekvizitlərinə (sənədin adı, yaradılma tarixi, ölçüsü, müəllifi və s.) görə axtarılması prosesidir. Bura kitabxana-informasiya sistemlərində istifadə olunan elektron kataloqları aid etmək olar.
- ***Təsvirə görə axtarış*** – məzmununa görə təsvirlərin axtarılması prosesidir. Axtarış sistemi təsvirlərin məzmunlarını tanıyır. Axtarış nəticəsində verilmiş təsvirə oxşar şəkillər istifadəçiyə təqdim edilir.
- ***Ünvanlı axtarış*** – axtarış sorğusunda göstərilən formal əlamətlərə görə sənədlərin axtarılması prosesidir. Burada sənədin ünvanı dedikdə onun yaddaş qurğusunda, informasiya saxlanclarında yerləşməsi qaydaları nəzərdə tutulur. Ünvan qismində web-serverin və ya web-səhifənin URL-ünvanı, bibliografik məlumat bazasında yazıların elementləri, sənədlərin saxlanclarda saxlanması ünvanları və s. çıxış edə bilər.

- **Semantik axtarış** – sənədlərin məzmununa görə onların axtarılması prosesidir. Burada sənədin və sorğunun məzmunu təbii dildən informasiya-axtarış dilinə çevrilir və onların axtarış obrazları yaradılır, axtarışın əlavə şərtləri göstərilən axtarış təsviri tərtib olunur.
- **Sənədli axtarış** – istifadəçinin sorğusuna uyğun olaraq, informasiya-axtarış sisteminin saxlancında ilkin sənədlərin və ya verilənlər bazalarında ikinci dərəcəli sənədlərin axtarılması prosesidir. Sənədli axtarışın iki növü vardır: ilkin sənədlərin axtarışına yönəlmiş kitabxana axtarışı və sənədlər haqqında biblioqrafik yazılar şəklində təqdim olunmuş məlumatların axtarışına yönəlmiş biblioqrafik axtarış.
- **Faktoqrafik axtarış** – axtarış sorğusuna uyğun olan faktların axtarılması prosesidir.

İnformasiya axtarışı sahəsində əsas anlayışlardan biri də **sorğu** anlayışıdır. Praktikada çox vaxt **axtarış sorğusu** və ya **informasiya sorğusu** terminlərindən istifadə olunur.

Burada **informasiya (axtarış) sorğusu** dedikdə istifadəçinin informasiyaya olan tələbatının informasiya-axtarış dilində ifadə olunan xüsusi forması başa düşülür. Bəzi axtarış sistemləri sorğunu təbii dildə daxil etməyə imkan verir.

İnformasiya axtarışının həyata keçirilməsi üçün informasiya-axtarış sistemləri reallaşdırılır. **İnformasiya-axtarış sistemi** (İAS) – istifadəçiləri maraqlandıran informasiyanın saxlanması, axtarılması və təqdim edilməsi üçün nəzərdə tutulmuş proqram-texniki sistemdir.

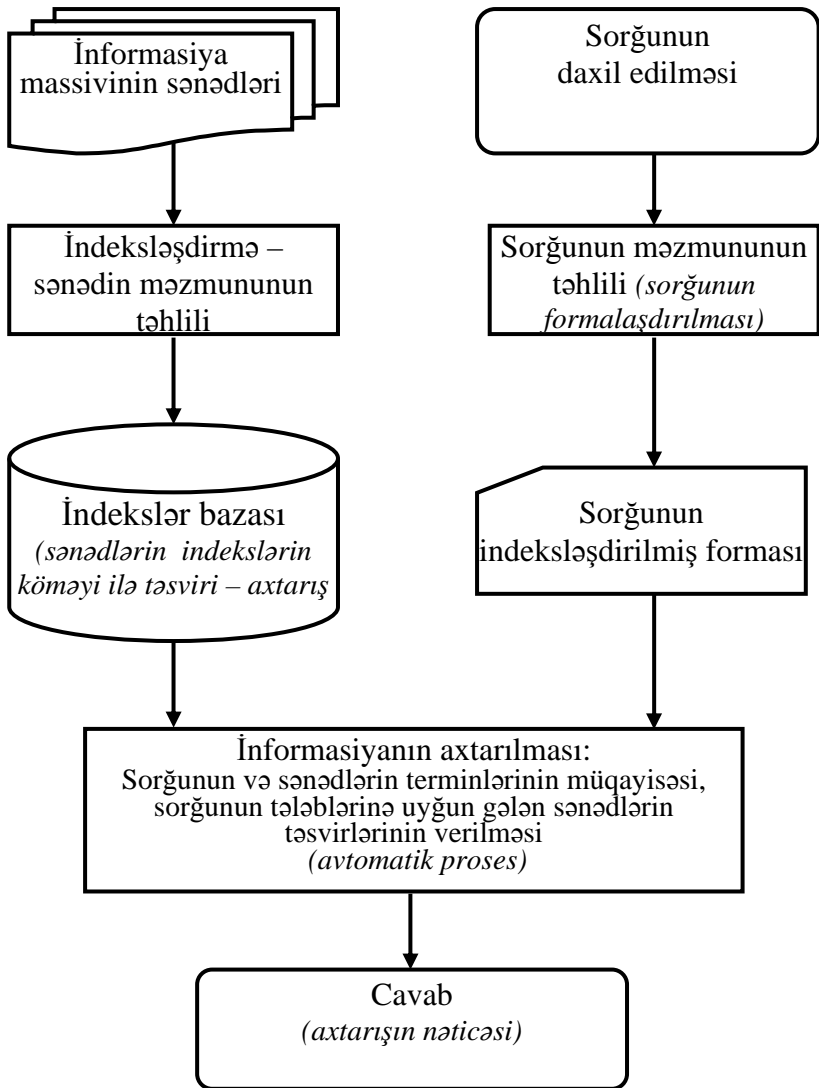
Bir qayda olaraq, İAS tərəfindən informasiya axtarışın təşkili üçün mümkün sənədlərin göstəriciləri (təsvirediciləri) çoxluğu yaradılır. Göstəricilər açar sözlər və ya söz

birləşmələrindən, onların aid olduqları sənədə istinaddan ibarət ola bilər. Açar sözlər və ya söz birləşmələri həmin sənədin hər hansı xassəsini daşıyır, onun məzmununa uyğun olur, onu mahiyyətə təsvir edir, istinadlar isə məhz bu xassələrə malik olan sənədlərə istiqaməti müəyyən edirlər. Ənənəvi sənədli İS-də göstəricilərin bir neçə formasından istifadə olunur. Daha geniş yayılmış göstəricilər müəllif, mövzu və sistemli göstəricilərdir.

Ayrı-ayrı söz və ya söz birləşmələrindən ibarət olan belə göstəricilərə (terminlərə, açar sözlərə və s.) *indekslər* deyilir. Sənədlərin göstəricilərinin, yəni onların indekslərinin yaradılması prosesi *indeksləşdirmə* adlanır. Sənədlərin indeksləşdirilməsi zamanı onlardan seçilib götürülən terminlər və ya açar sözlər *indeksləşdirmə terminləri* adlanır.

Adətən, hər bir indeksin aid olduğu sənəd üçün təxmin edilən *vaciblik dərəcəsini* əks etdirən *çəki əmsalları* ayrı-ayrılıqda hesablanır və ona əlavə edilərək bazada saxlanılır. İndeksləşdirmə prosesində istifadə olunan terminlər çoxluğuna *lüğət*, informasiya ehtiyatlarının indeksləşdirilməsi nəticəsində alınmış göstəricilər (indekslər) çoxluğuna *indekslər bazası*, bəzən isə sadəcə *indeks* deyilir. Ayrı-ayrı sənədlərin məzmununu təsvir edən indekslər (göstəricilər) çoxluğu *sənədin axtarış obrazı* adlanır.

İndekslər bazası yaradıldıqdan sonra İAS axtarış maşınları adlanan xüsusi proqramların köməyi ilə istifadəçilərə tərtib etdikləri sorğular vasitəsilə həmin bazalara müraciət etmək imkanını verir. Ümumiyyətlə, axtarış prosesi istifadəçinin sorğusu ilə sənədin indekslər bazasında olan axtarış obrazının müqayisə edilməsindən ibarətdir. Ona görə də sistemə daxil olmuş sorğular da informasiya massivinin sənədlərinə analogi olaraq indeksləşdirilməlidir (şəkl.4.2).



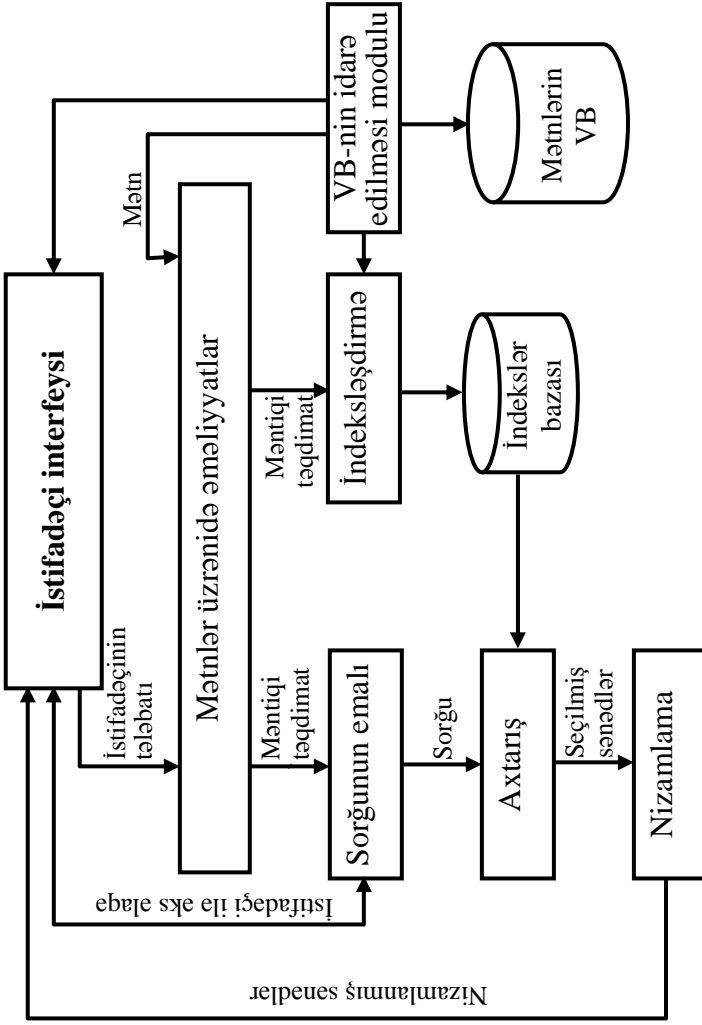
Şək.4.2. İnformasiya axtarışının sadələşdirilmiş prosedurası (indeksləşdirmə və axtarış)

Sənədlərin istifadəçinin sorğusuna relevant olub-olmaması barəsində qərar həmin sənədlərə və sorğuya aid olan terminlər çoxluqlarının müqayisəsi nəticəsində qəbul edilir. Nəticədə, istifadəçilərə elə sənədlər təqdim olunur ki, onların axtarış obrazları istifadəçinin sorğusunun axtarış obrazı ilə müəyyən olunmuş dərəcədə üst-üstə düşmüş olsun. Qeyd olunmalıdır ki, burada çox vaxt sorğunun və sənədlərin axtarış obrazları qismən üst-üstə düşür. Başqa sözlə, bu obrazlar tam üst-üstə düşmür, lakin onlar çox yaxın olurlar.

Axtarış nəticəsində sənədlərin sorğuya uyğun olub-olmamasının müəyyənləşdirilməsi kriterisi hər bir sənəd və sorğu cütünü üçün oxşarlıq (relevantlıq) əmsalının hesablanması prinsipinə əsaslanı bilər. Bu halda oxşarlıq əmsalının qiyməti əvvəlcədən müəyyən edilmiş həddən (sərhəd qiymətindən) yuxarı olan sənədlər nəticə (relevant sənədlər) qismində istifadəçiyə verilir. Tapılmış sənədlər sorğuya oxşarlıq dərəcələrinə görə nizamlanır və bu qiymətlərin azalması ardıcılığı ilə istifadəçiyə təqdim olunur. İnformasiya axtarışı prosesinin ümumi strukturu 4.3 sayılı şəkildə verilmişdir.

Ümumiyyətlə, informasiya axtarışı proqram-texniki üsul və vasitələrdən istifadə etməklə dörd mərhələdə həyata keçirilir:

- Axtarışın məqsəd və vəzifələrinin formalaşdırılması – fərd və ya təşkilat qarşısında duran vəzifələrin icrası zamanı yaranan informasiya tələbatından irəli gələrək həyata keçirilir. Axtarışın vəzifələrinin müəyyənləşdirilməsi axtarılan informasiyanın (ədəbiyyatın) sahəsini və xüsusiyyətlərini müəyyən etməyə, axtarışın həyata keçirilməsi üçün plan və proqramın tərtib edilməsini asanlaşdırmağa imkan verir.



Şək.4.3. İnformasiya axtarışı prosesinin strukturu

- Axtarışın işçi proqramının işlənib hazırlanması – axtarış qarşısında qoyulmuş vəzifələri maksimal dərəcədə açılması və konkretləşdirilməsini nəzərdə tutur. Burada axtarışın obyektı, növü, üsulları, mümkün istiqamətlər, tematik, xronoloji, dil, janr məhdudiyətləri, obyektiv və subyektiv çətinliklər, dolğunluq dərəcəsi, nəticələrin təqdim olunması formaları müəyyən edilir.
- Axtarışın reallaşdırılması – axtarış əvvəlcə mövzuya və məqsədə uyğun şəkildə aparılır, informasiya mənbələrini geniş əhatə etmək zərurəti yarandıqda isə bibliografik və sənədli axtarış üsullarından istifadə olunur. Sonra informasiya axtarışı konkretləşdirilir, bilavasitə həll edilən məsələnin məzmunu ilə məhdudlaşdırılır.
- Axtarışın nəticələrinin tərtib edilməsi – burada axtarış nəticəsində tapılmış informasiya mənbələrinin hansı şəkildə istifadəçiyə təqdim ediləcəyi müəyyən edilir.

4.2. İnformasiya-axtarış dili

Qeyd olunduğu kimi, İAS-a müraciət zamanı istifadəçilərə sistemin sorğular dilini reallaşdıran interfeys xidməti təqdim olunur. İstifadəçinin informasiya tələbatı məhz İAS-ın təqdim etdiyi vasitələrin köməyi ilə onun başa düşdüyü şəkildə ifadə olunmalıdır. Başqa sözlə, axtarış sorğusu İAS-ın formal daxili sorğular dilində formalaşdırılmalıdır.

Adətən, İAS-da reallaşdırılan sorğular dili müəyyən həddə qədər süni dil olur. Belə ki, sistemlə ünsiyyətin yaradılması, sorğuların formalaşdırılması, daxil edilməsi və informasiyanın təqdim olunması üçün əsas vasitə kimi təbii dilin tam istifadəsinin mümkün olmaması İAS-da in-

formasiya-axtarış dili adlanan süni dil vasitələrinin tətbiq olunması zərurətini yaratmışdır.

İnformasiya-axtarış dili (İAD) – saxlanılması və axtarışın aparılması məqsədilə sənədlərin və axtarış sorğularının məzmunlarının indeksləşdirilməsi, eləcə də axtarış sorğularının tərtib olunması qaydalarını müəyyən edən formallaşdırılmış süni dildir.

Praktikada İAD-ın bir neçə növünə rast gəlinir.

Sənədli İAD – saxlanılması və axtarışın aparılması məqsədilə sənədlərin (onların hissələrinin) məzmunlarının indeksləşdirilməsi üçün nəzərdə tutulmuş informasiya-axtarış dilidir.

Təsnifatlandırıcı İAD – sənədlərin və informasiya sorğularının seçilmiş təsnifat sisteminin (sənədlərin təsnifatının) anlayışları və kodları vasitəsilə indeksləşdirilməsi üçün nəzərdə tutulmuş informasiya-axtarış dilidir. Təsnifatlandırıcı İAD-ın lüğət tərkibinə əvvəlcədən həm sadə, həm də mürəkkəb terminlər (söz birləşmələri, ifadələr, terminlərin başqa dillərə tərcüməsi, sinonimləri, assosiativ sözləri, qohum sözləri və s.) çoxluğu daxil edilir. Sənədlərin və sorğuların məna və məzmunlarının təqdim edilməsi üçün bu çoxluğun ayrı-ayrı elementləri, eləcə də hazır mürəkkəb ifadələri istifadə olunur. Belə dillərdə mürəkkəb sorğuların tərtib edilməsi prosesi mövcud lüğət tərkibindən uyğun gələn mürəkkəb elementlərin seçilməsi prosedurası ilə əvəz edilir. Beləliklə, sənədlər əvvəlcədən dilin tərkibində mövcud olan sadə və mürəkkəb leksik elementlərin köməyi ilə sistemdə müəyyən olunmuş ayrı-ayrı təsnifat siniflərinə (məsələn, tematik rubrikatorun ayrı-ayrı profillərinə) aid edilir.

Təsviredici İAD – sənədlərin və informasiya sorğularının təsviredicilər (deskriptorlar, açar sözlər) vasitəsilə

koordinata görə indeksləşdirilməsi üçün nəzərdə tutulmuş informasiya-axtarış dilidir. Təsviredici İAD sorğuların tərtib olunması zamanı leksik vahidlərin daxil edilməsi və birləşdirilməsi yolu ilə axtarılan sənədlərin məna və məzmunlarının bu sorğularda təsvir olunmasına imkan verir. Belə dillərdə əvvəlcədən hazırlanmış cümlələr və ifadələr mövcud olmur, mürəkkəb sorğuların tərtib olunması zamanı onların forma və strukturlarına hər hansı məhdudiyyət qoyulmur.

Təsviredici informasiya-axtarış dillərini leksik tərkiblərinə görə də iki yerə ayırırlar: nəzarət olunan leksikalı İAD və sərbəst leksikalı İAD. *Nəzarət olunan leksikalı dillərin* leksik tərkibi ciddi məhdudlaşdırılmış və təsbit edilmiş olur, *sərbəst leksikalı dillərin* leksik tərkibinə isə məhdudiyyət qoyulmur və onlar sistemin fəaliyyəti prosesində yeni leksik vahidlərin daxil edilməsi yolu ilə daim təkmilləşdirilir.

Nəzarət olunan lüğətlərdə terminlər bazaya sistemin administratoru tərəfindən (və ya onun icazəsi ilə operator tərəfindən) əlavə oluna bilər. Bütün yeni sənədlər yalnız bu bazada olan terminlər əsasında indeksləşdirilir. Sərbəst lüğətlərdə sənədlərdə yeni rast gəlinən terminlər avtomatik olaraq bazaya əlavə edilir. Sərbəst lüğət axtarış prosesində aktuallaşdırılan zaman təsbit edilmiş olur. Belə bazalarda müəyyən dəyişikliklər aparıldıqdan sonra yeniliklərin aktuallaşması üçün bazanın yenidən yüklənməsi tələb olunur.

Sintaktik konstruksiyaların formalaşdırılması üçün istifadə olunan qaydalarına görə təsviredici İAD iki yerə bölünür: qrammatikalı və qrammatikasız. *Qrammatikalı İAD-da* sənədlərin təsviri və sorğuların formalaşdırılması üçün ciddi qaydalar istifadə olunur. Təbii dildən fərqli ola-

raq, belə dillərdə ciddi təqdimat formaları istifadə oluna bilər. Məsələn, təbii dildə olan

*"axtarış sistemi sorğular dilinə malikdir",
sintaksis: {subyekt, obyekt, hərəkət}*

cümləsi aşağıdakı şəkildə təsvir oluna bilər:

*"malik olmaq, axtarış sistemi, sorğular dili",
sintaksis: {hərəkət, subyekt, obyekt}.*

Qrammatikasız İAD-da dillərində isə heç bir məhdudiyət və sintaktik qaydaları olmur.

Faktoqrafik İAD – faktoqrafik informasiya massivlərində faktların və informasiya sorğularının indeksləşdirilməsi üçün nəzərdə tutulmuş informasiya-axtarış dilidir.

Obyekt-əlamətli İAD – faktların təsvirinin obyektlərin onlara aid olan əlamətlər və bu əlamətlərin qiymətləri göstərilməklə tərtib edilmiş siyahı şəklində indeksləşdirilməsi üçün nəzərdə tutulmuş faktoqrafik informasiya-axtarış dilidir.

Verbal İAD – təbii dilin leksik vahidlərinin təqdim edilməsi üçün söz və ifadələri onların orfoqrafik formalarında istifadə edən informasiya-axtarış dilidir.

Təbii ki, istifadə zamanı İAD həm informasiya-axtarış sistemləri, həm də istifadəçilər üçün aydın, rahat və əlverişli olmalıdır. Bu, o deməkdir ki, informasiya-axtarış dilinin işlənilməsi zamanı dilin forması, lüğət tərkibi və sintaksisi yaxşı və hərtərəfli düşünülməlidir.

Dilin lüğət tərkibi – axtarış sisteminin dilində mümkün sənədlərin və sorğuların təsviri və təqdim olunması üçün istifadə edilən deskriptorlardan, yəni terminlərdən və açar sözlərdən, eləcə də tezaurusdan, sinonimlər və assosiasiya

yaradan (assosiativ) sözlər çoxluqlarından ibarət kompleksdir.

İnformasiya-axtarış tezaurusu – təsviredici İAD-ın leksik vahidlərinin paradiqmatik münasibətlərini (sözlərin hallanması və formalarının dəyişməsi qaydalarını) təsbit edən lüğətdir. **Tezaurus** – dilin sözlərinin (onların mətnlərdə istifadəsi nümunələri ilə birlikdə) maksimal dərəcədə tam təqdim edildiyi lüğətdir. **Tezaurus** – hər hansı bilik sahəsinə aid olan sözlərin mövzu prinsipinə uyğun düzül-düyü və leksik vahidlər arasındakı semantik münasibətlərin göstərildiyi lüğətdir. Tezaurus indeksləşdirmə üçün tövsiyə olunan sözlərin (deskriptorların) seçilməsi yolu ilə sənədlərin məzmununun təsvir edilməsi üçün istifadə olunan açar sözlərin siyahısını özündə saxlayır.

Dilin sintaksisi – sistemin dilinin lüğət tərkibinin elementlərindən mənalı (məniyyəti) onun baza lüğətinin ayrı-ayrı sözləri vasitəsi ilə ifadə edilməsi mümkün olmayan leksik vahidlərin yaradılması qaydaları çoxluğudur.

Leksik vahidlər dedikdə bir sənədi və ya sənədlər qrupunu, eləcə də sorğuları digərləri ilə müqayisə etməyə imkan verən açar sözlər, terminlər, söz birləşmələri və s. başa düşülür.

İAD-ın qrammatikası – axtarış obrazlarının və axtarış göstərişlərinin informasiya-axtarış dilinin leksik vahidlərindən tərtib edilməsi qaydalarıdır.

İAD vasitəsilə sənədlərin və sorğuların təsvir edilməsi prinsiplərinə və qaydalarına **indeksləşdirmə qaydaları** deyilir, axtarış sorğusuna uyğun sənədlərin tapılması mexanizmlərinə isə **axtarış qaydaları** deyilir.

Yuxarıda qeyd olunanlara analoji olaraq, internetin axtarış sistemlərinin İAD ilə bağlı olan və onların qurulmasında nəzərə alınan əsas anlayışlar aşağıdakılardır.

1. *İndeksləşdirmə terminləri lüğəti* – deskriptorlar (anlayışların adları), açar sözlər və s. şəklində axtarış terminləri çoxluğudur. Bu terminlər söz birləşmələri şəklində sorğulara və ya indekslərə daxil edilə bilər, məsələn, "informasiya axtarışı".
2. *Məntiqi operatorlar*. Sorğuların formalaşdırılması üçün informasiya-axtarış dillərində "AND" ("VƏ"), "OR" ("VƏ YA"), "NOT" ("YOX") və s. kimi məntiqi operatorlar, yəni bu əməlləri istifadə olunur. Bu operatorların köməyi ilə qurulmuş məntiqi operatorların sintaksisi müxtəlif sistemlərdə fərqli ola bilər.
3. *Sorğuların normallaşdırılması*. Eyni terminlərin və açar sözlərin vahid formada istifadə olunması məqsədilə informasiya-axtarış sistemlərində leksik vahidlərin normallaşdırılması (vahid formaya gətirilməsi) mexanizmləri reallaşdırılır. Sənədlərin və sorğuların leksik vahidlərinin normallaşdırılması prinsipi sorğu formalaşdırılan zaman axtarış dilinin imkanlarının köməyi ilə istifadəçi tərəfindən müəyyənləşdirilə və ya bilavasitə sistemin axtarış alqoritmlərində reallaşdırıla bilər. İstifadəçilər sistemin axtarış alqoritmlərində reallaşdırılan normallaşdırma mexanizmlərinin olması barədə təsəvvürə malik olmur və axtarış prosesində bu mexanizmlərin mövcudluğunu hiss etmirlər. Qeyd olunmalıdır ki, sorğuların normallaşdırılması üçün başqa mexanizmlərdən də istifadə oluna bilər.

4. *Dilin qrammatikası*. Müəyyən anlayışların mənalarının genişlənməsi və daralması üçün indeksləşdirmə terminlərinin istifadəsi ilə yanaşı informasiya-axtarış dilinin tərkibinə digər mexanizmlər də daxil olur. Onlara aşağıdakıları aid etmək olar:
- söz birləşmələrinin istifadəsi;
 - bir neçə terminin birlikdə sənədlərə daxil olmasını göstərən "AND" ("VƏ") məntiqi operatorunun tətbiqi;
 - sənədlər terminlər arasındakı məsafəni göstərən "NEAR" ("YANINDA") məntiqi operatorunun reallaşdırılması və s.
5. *Relevantlıq (qeyri-relevantlıq) kriterisi*. Dilin tərkibində informasiya massivində olan sənədlərin axtarış sorğusuna relevant və qeyri-relevant olmasının təsdiq edilməsi, eləcə də bu sənədlərin axtarışının nəticəsi kimi istifadəçiyə verilməsi və ya verilməməsi qaydaları müəyyən edilir. Axtarış nəticəsində seçilmiş sənədlər istifadəçiyə nizamlanmış siyahı şəklində təqdim olunur. Bunun üçün sənədin istifadəçinin sorğusuna yaxınlıq dərəcəsini (qiymətini) müəyyən edən oxşarlıq əmsallarından istifadə olunur. Oxşarlıq əmsalları sorğudakı açar sözlərin sənədlər üçün termin qismində əhəmiyyətini göstərən çəki əmsallarının köməyi ilə hesablanır.
6. *Axtarışın əlavə şərtləri*. Axtarışın nəticəsinin dəqiqliyinin və dolğunluğunun yüksəldilməsi üçün müxtəlif əlavə axtarış şərtləri istifadə oluna bilər. Bu şərtlərə sorğuda göstərilmiş terminlərin sənədin müəyyən olunmuş hissələrində (adında, başlığında, girişində, xülasəsində və s.) axtarılması, axtarış sahəsinin da-

raldılması, sənədlərin yaradılması tarixlərinə qoyulmuş zaman intervallarının, müxtəlif lüğətlərin (tezaurusların, assosiativ sözlərin, sinonimlərin və s.) istifadəsi kimi şərtlər aid edilə bilər.

7. *Axtarışın nəticələrinin verilməsi forması.* İstifadə üçün rahatlığı təmin etmək məqsədilə tapılmış relevant sənədlərin siyahısı istifadəçilərə xüsusi formalarda təqdim olunur. Bu siyahı, əsasən, aşağıdakı parametrlərlə müəyyən olunur:

- tapılan və istifadəçilərə verilən sənədlərin siyahısının nizamlanması (daha vacibdən nisbətən az vacibə doğru sıraya düzülməsi);
- siyahının forması;
- siyahıya sənədlər haqqında əlavə məlumatların (məsələn, mətn fraqmentinin, sorğunun sözlərinin sənəddə rastgəlinmə tezliyi, sənədin sorğuya yaxınlıq dərəcəsi və s.) daxil edilməsi;
- sənədlərin siyahısının istifadəçiyə verilməsi porsiyasının həcmi, yəni interfeys pəncərəsinin çıxış səhifələrində sənədlərin sayı;
- axtarış nəticəsində istifadəçiyə təqdim olunan sənədlərin ümumi sayına qoyulan məhdudiyət.

8. *İstifadəçi ilə əks əlaqə.* Axtarış prosesində istifadəçiyə öz sorğularının dəqiqləşdirilməsi, sənədlərin təqdimatının təkmilləşdirilməsi, eləcə də əvvəldən tapılmış sənədlər arasından göstərilmiş konkret sənədə "oxşar" olan sənədlərin axtarışının aparılması imkanları verilir.

Nəhayət, qeyd olunmalıdır ki, informasiya-axtarış sistemlərində istifadəçi ilə sistem arasında yaradılan intellektual dialoq istifadəçiyə öz sorğularını daha dəqiq və rahat

şəkildə formalaşdırmaq imkanı verir. Bu işə yekunda axtarış prosesinə və onun nəticələrinə təsir edir.

İstifadəçi sorğusuna daxil edilən açar sözlərin qabaqcadan indeksləşdirilmiş sənədlərin axtarış obrazları ilə müqayisəsindən sonra axtarış sistemi informasiya-axtarış dilinin köməyi ilə istifadəçiyə konkret sənədlərə istinadların siyahısını təqdim edir. İAS-a görə istifadəçiyə verilən sənədlər onun sorğusuna ən relevant sənədlərdir.

İnternetdə fəaliyyət göstərən Google, AltaVista, Excite, Open Text, Yahoo, Yandex, Rambler və s. kimi axtarış sistemlərində ən sadədən ən mürəkkəbə qədər müxtəlif sorğuları formalaşdırmağa imkan verən güclü İAD reallaşdırılmışdır.

İntellektual istifadəçi interfeysinin və sorğuların formalaşdırılması xidmətinin təşkili üçün istifadə olunan İAD sənədlərin təqdim olunması və informasiyanın saxlanması modellərindən, eləcə də onların indeksləşdirilməsi və axtarılması üsullarından birbaşa asılıdır.

4.3. İnformasiya axtarışının nəticəsinin qiymətləndirilməsinin əsas meyarları

İAS tərəfindən tapılmış sənədlərin axtarış sorğusuna uyğunluq dərəcəsini qiymətləndirmək üçün bir çox üsullar mövcuddur. Tapılmış sənədin sorğuya uyğunluğu dərəcəsi anlayışı subyektiv anlayış olub, sorğunu tərtib edən və axtarışın nəticəsinə qiymətləndirən adamdan asılı olur.

İnformasiya axtarışının nəticəsinin təhlili və qiymətləndirilməsi üçün iki anlayışdan – relevantlıq və pertinentlik anlayışlarından istifadə olunur.

Relevantlıq – axtarış nəticəsində tapılaraq istifadəçiyə təqdim edilmiş sənədlərin məzmununun informasiya sorğusunun məzmununa uyğunluğu dərəcəsidir. Məzmunları istifadəçinin sorğusuna cavab verən sənədlər **relevant**, cavab verməyənlər isə **qeyri-relevant sənədlər** adlanır.

Relevantlıq anlayışının iki formasını fərqləndirirlər:

- formal relevantlıq;
- faktiki relevantlıq.

Formal relevantlıq – axtarış mexanizmlərinin köməyi ilə İAS tərəfindən istifadəçinin axtarış sorğusunun informasiya massivində olan sənədlərlə tutuşdurulması (müqayisə edilməsi) yolu ilə müəyyən olunan relevantlıqdır.

Faktiki relevantlıq dedikdə axtarış nəticəsində İAS-dan alınmış sənədlərin məzmununun istifadəçi tərəfindən araşdırılması (dərək edilməsi) və onun tələbatına yaxınlıq dərəcəsinin təyin edilməsi yolu ilə müəyyən olunan relevantlıq başa düşülür.

Hər bir istifadəçi özünün informasiyaya olan tələbatını dəqiq bilməsinə baxmayaraq, obyektiv səbəblərdən bu tələbatı axtarış sorğusunda qeyri-adekvat şəkildə formalaşdırma və düzgün ifadə etməyə bilər. Belə olduqda istifadəçinin konkret sorğusuna cavab olaraq İAS-ın verdiyi informasiya ehtiyatlarının müəyyən hissəsi (bəzən əksəriyyəti) həqiqətdə istifadəçinin tələblərinə (ehtiyacına) uyğun olmur. Ona görə də axtarışın nəticəsinin keyfiyyətinin müəyyənləşdirilməsi üçün İAS-ın verdiyi sənədlərin istifadəçinin sorğusuna deyil, onun informasiya tələbatına uyğunluğunun qiymətləndirilməsi böyük əhəmiyyət kəsb edir. Bu məqsədlə pretinentlik anlayışından istifadə olunur.

Pertinentlik – informasiya axtarışı nəticəsində tapılmış sənədlərin məzmununun istifadəçinin axtarış sorğusu şəklində formalaşdırılmış informasiya tələbatına uyğunluğu

dərəcəsidir. Məzmunları istifadəçinin informasiya tələbatına cavab verən sənədlər *pertinent*, cavab verməyənlər isə *qeyri-pertinent sənədlər* adlanır.

Qeyd etmək lazımdır ki, pertinentlik anlayışı mahiyyət etibarı ilə faktiki relevantlıq anlayışına yaxındır. Aydınır ki, tapılmış sənədlərin relevantlıq dərəcələri axtarışın keyfiyyətindən, yəni İAS-ın "intellektual qabiliyyətindən", pertinentlik dərəcəsi isə istifadəçidən, onun mövzu sahəsini bilməsi səviyyəsindən, öz tələbatını ifadə və təsvir etmək bacarığından asılı olur. Təbii ki, burada İAS-ın axtarış aparatı, eləcə də informasiya-axtarış dili böyük rol oynayır.

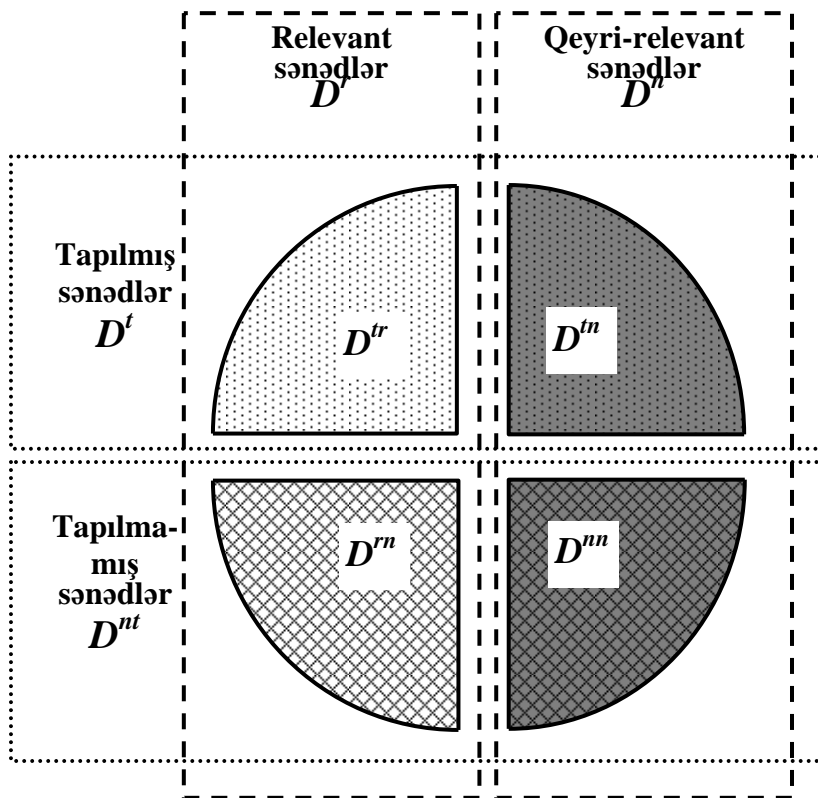
İAS tərəfindən tapılan relevant sənədlər relevantlıq dərəcəsinə görə nizamlanmış, yəni sıraya düzülmüş siyahı şəklində istifadəçiyə təqdim edilir. Siyahının əvvəlində daha relevant sənədlər, sonra isə nisbətən az relevant olanlar və s. gəlir.

Praktikada informasiya axtarışının nəticəsinin keyfiyyəti və İAS-ın işinin səmərəliliyi əsas iki qarşılıqlı əlaqəli parametrlə xarakterizə olunur:

- axtarışın dəqiqliyi;
- axtarışın dolğunluğu.

Bu anlayışları şərh etmək üçün aşağıdakı işarələmələri qəbul edək. Tutaq ki, D – informasiya massivində olan sənədlərin sayı, D^r – informasiya massivində olan bütün relevant sənədlərin sayı, D^n – informasiya massivində olan bütün qeyri-relevant sənədlərin sayı, D^t – İAS tərəfindən relevant sənəd qismində tapılıb istifadəçiyə verilmiş sənədlərin ümumi sayı, $D^{n'}$ – informasiya massivində olan, lakin İAS tərəfindən tapılıb istifadəçiyə verilməmiş sənədlərin ümumi sayı, $D^{r'}$ – İAS tərəfindən tapılıb verilmiş və istifadəçinin sorğusuna həqiqətən relevant olan sənədlərin sayı, $D^{n''}$ – İAS tərəfindən tapılıb verilmiş, lakin sorğuya

relevant olmayan sənədlərin sayı, D^m – informasiya massivində sorğuya relevant olan, lakin İAS tərəfindən tapılmayan sənədlərin sayı, D^{mn} – informasiya massivində sorğuya relevant olmayan və İAS tərəfindən tapılmayan sənədlərin sayıdır (şək.4.4).



Şək.4.4. İnformasiya axtarışının əsas meyarları

Aydındır ki,

$$D = D^r + D^n = D^t + D^{nt}$$

$$D^r = D^{tr} + D^{rn}$$

$$D^n = D^{tn} + D^{nn}$$

$$D^t = D^{tr} + D^{tn}$$

$$D^{nt} = D^{rn} + D^{nn}$$

ifadələri doğru olar.

Axtarışın dəqiqliyi İAS tərəfindən tapılıb verilmiş sənədlərin ümumi sayının bu sənədlər arasında istifadəçinin sorğusuna həqiqətən relevant olanlarının sayına tərs nisbəti ilə xarakterizə olunur.

$$P = \frac{D^{tr}}{D^t} = \frac{D^{tr}}{D^{tr} + D^{tn}}, \quad (4.1)$$

burada P (“precision” sözünün baş hərfi) – axtarışın dəqiqliyidir.

Axtarışın dolğunluğu İAS tərəfindən tapılıb verilmiş və istifadəçinin sorğusuna həqiqətən relevant sənədlərin sayının informasiya massivində olan bütün relevant sənədlərin (tapılan və tapılmayan) sayına nisbəti ilə müəyyən olunur.

$$R = \frac{D^{tr}}{D^r} = \frac{D^{tr}}{D^{tr} + D^{rn}}, \quad (4.2)$$

burada R (“recall” sözünün baş hərfi) – axtarışın dolğunluğudur.

Başqa sözlə, dolğunluq və dəqiqlik terminlərini aşağıdakı kimi izah etmək olar: *dəqiqlik* – İAS-ın relevant olmayan sənədləri axtarışın nəticəsinə daxil etməməsi, yəni istifadəçiyə verməməsi, *dolğunluq* isə İAS-ın informasiya massivində olan bütün mümkün relevant sənədləri tapıb istifadəçiyə təqdim etməsi qabiliyyətidir.

İAS-ın işinin nəticəliliyinin qiymətləndirilməsi üçün “seçim qabiliyyəti” və “küy” adlanan parametrlər də böyük əhəmiyyət kəsb edir.

Seçim (yığım) qabiliyyəti İAS tərəfindən tapılıb verilmiş qeyri-relevant sənədlərin sayının axtarış aparılan informasiya massivində olan bütün qeyri-relevant sənədlərin sayına nisbəti ilə müəyyən olunur.

$$E = \frac{D^{nt}}{D^n} = \frac{D^{nt}}{D^{tn} + D^{nm}}, \quad (4.3)$$

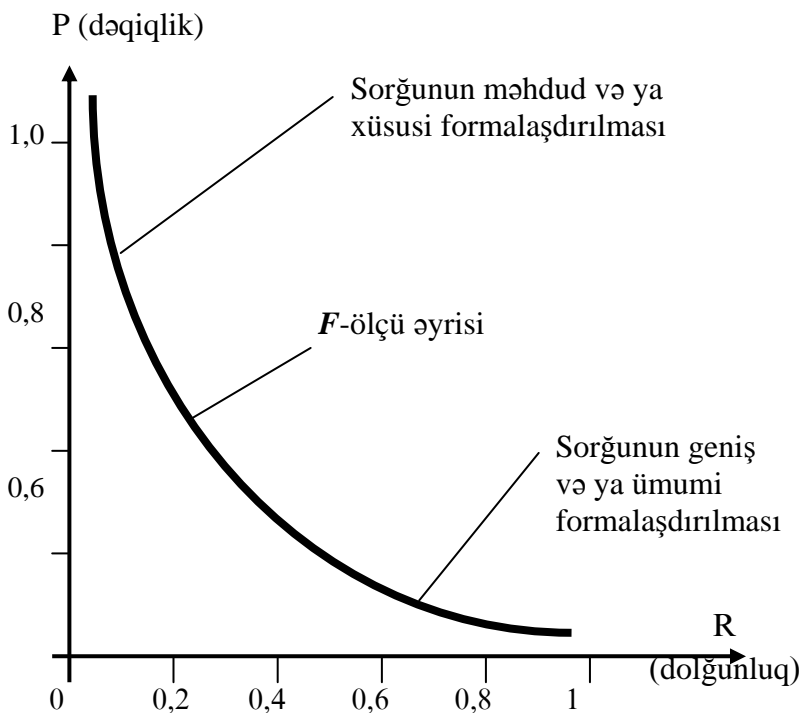
burada E (“election” sözündən) – axtarışın seçim qabiliyyətini göstərir.

Axtarışın dəqiqliyinin əksi olan *küy* parametri İAS tərəfindən tapılıb verilmiş sənədlərin ümumi sayının bu sənədlər arasında istifadəçinin sorğusuna qeyri-relevant olanlarının sayına tərs nisbəti ilə xarakterizə olunur və axtarışın nəticəsində qeyri-relevant sənədlərin olması dərəcəsini göstərir.

$$N = 1 - P = 1 - \frac{D^{tr}}{D^t} = \frac{D^{tn}}{D^{tr} + D^{tn}}, \quad (4.4)$$

burada N (“noise” sözünün baş hərfi) – axtarışın “küy” parametridir.

Axtarışın dəqiqliyi və dolğunluğu parametrləri 0-dan (ən pis hal, sənəd sorğuya uyğun deyil, yəni onların uyğunluq dərəcəsi sıfır faizdir) 1-ə (ən yaxşı hal, yəni sənəd sorğuya 100% uyğun gəlir) kimi intervalda qiymətlər alır. Onlar bir-biri ilə bilavasitə qarşılıqlı əlaqəyə malikdirlər. Adətən, dolğunluğu yüksəltmək üçün həyata keçirilən bütün tədbirlər dəqiqliyin azalmasına və əksinə, dəqiqliyin yüksəldilməsi üsulları dolğunluğun azalmasına gətirib çıxarır (şək.4.5).



Şək.4.5. Dolğunluq və dəqiqlik arasında asılılıq

Dəqiqlik və dolğunluq arasında harmonik orta qiyməti müəyyənləşdirmək üçün *F-ölçü* adlanan parametrdən istifadə olunur.

$$F = \frac{2PR}{P + R} \quad (4.5)$$

Praktikada ayrı-ayrı hallarda müxtəlif istifadəçilər tərəfindən bu parametrlərə, yəni İAS-ın xarakteristikalarına qoyulan tələblər müxtəlif olur. Bəzi istifadəçilər yüksək dolğunluğun təmin olunmasını arzu edirlər. Belə ki, onlar üçün mövzu sahəsinə müəyyən aidiyyəti olan hər hansı dərəcədə maraq kəsb edə biləcək bütün sənədlərin əldə olunması zəruridir. Bəzən isə istifadəçilər yüksək dəqiqliyin təmin olunmasını, yəni onların istəklərinə uyğun olmayan bütün sənədlərin atılmasını və axtarışın nəticəsi qismində onlara təqdim olunmamasını tələb edirlər.

Ən səmərəli İAS dolğunluq və dəqiqlik əmsalları arasında optimal tarazlığı tapır və mümkün qədər həm yüksək dolğunluğu, həm də yüksək dəqiqliyi təmin etməyə çalışır. Dolğunluğun və dəqiqliyin optimal səviyyəsinin təmin edilməsi üçün İAS-da intellektual informasiya axtarışı və indeksləşdirmə üsulları reallaşdırılmalıdır.

Axtarış prosesində aşağıdakı şərtlərin (parametrlərin) dəqiqləşdirilməsi yolu ilə axtarış kriterilərini bu və ya digər dərəcədə idarə etmək mümkündür:

- axtarışın məqsədi;
- axtarışın mövzusu;
- axtarılan sənədin (ədəbiyyatın) növü;
- axtarış üsulu;
- xronoloji əhatə dairəsi;

- coğrafi əhatə dairəsi;
- mövzu (istehsal) sahəsi üzrə əhatə dairəsi.

Qeyd olunan parametrlərin konkretləşdirilməsi daha məqsədyönlü, hərtərəfli, dərin, rasionall və səmərəli informasiya axtarışını həyata keçirməyə imkan verir.

V FƏSİL

İNFORMASIYA AXTARIŞININ EFFEKTİVLİYİNİN YÜKSƏLDİLMƏSİ ÜSULLARI

**Axtarışın effektivliyinin qiymətləndirilməsinə
yanaşmalar**

Axtarışın dolğunluğunun yaxşılaşdırılması üsulları

Axtarışın dəqiqliyinin yaxşılaşdırılması üsulları

**Relevantlığın yüksəldilməsi üçün istifadəçi ilə əks
əlaqənin istifadəsi**

5.1. Axtarışın effektivliyinin qiymətləndirilməsinə yanaşmalar

Axtarışın effektivliyini tapılıb istifadəçiyə verilmiş sənədlərin keyfiyyət və kəmiyyət xarakteristikalarına əsasən qiymətləndirmək olar. *Keyfiyyət xarakteristikaları* dedikdə tapılıb istifadəçiyə verilmiş sənədlərin onun sorğusuna uyğunluq (relevantlıq) dərəcəsi başa düşülür. *Kəmiyyət xarakteristikaları* isə tapılıb istifadəçiyə verilmiş sənədlərin sayının informasiya massivində olan, lakin tapılmamış relevant sənədlərin sayına nisbəti ilə müəyyən olunur. Bu xarakteristika dolğunluq parametrinə analogiya təşkil edir.

Keyfiyyət xarakteristikalarının müəyyənləşdirilməsi üçün axtarışın nəticəsi ilə istifadəçinin tələbatının ödənilməsi dərəcəsini müqayisə etmək lazımdır. Kəmiyyət xarakteristikaları isə statistik məlumatların köməyi ilə hesablanır.

Qeyd olunmalıdır ki, hər iki halda bir sıra çətinliklər qarşıya çıxır. Belə ki, birinci halda istifadəçi ilə əks əlaqənin reallaşdırılması tələb olunur, lakin belə əlaqəni həmişə arzu olunan səviyyədə həyata keçirmək mümkün olmur. İkinci halda isə informasiya massivində olan, lakin tapılmamış relevant sənədlərin sayının müəyyənləşdirilməsi zəruridir.

İAS-ın effektivliyinin qiymətləndirilməsi üçün axtarışın əsas parametrlərinin – dəqiqliyin və dolğunluğun dəqiq müəyyənləşdirilməsi tələb olunur. Bunları isə sənədlərinin sayı təsbit olunmuş sistemlərdə həyata keçirmək mümkün olur. Belə sistemlərdə standart sorğular çoxluğu, eləcə də hər bir emal olunan sorğuya relevant və qeyri-relevant sənədlərin sonlu çoxluqları mövcud olur.

Real fəaliyyət göstərən axtarış sistemləri yalnız öz serverlərində olan sənədlərin deyil, həmçinin müxtəlif saytlarda yerləşən informasiya ehtiyatlarının indekslərini toplayır. Ona

görə də praktikada internetin informasiya fəzası, o cümlədən web-fəza üçün müvafiq ölçmələri aparmaq olduqca çətinidir.

İAS-ın effektivliyinin qiymətləndirilməsi zamanı müəyyən çətinliklərin rast gəlinməsinə baxmayaraq, bütövlükdə axtarışın effektivliyinin artırılması üçün müxtəlif yanaşmalara əsaslanan üsullar mövcuddur. Onlar arasından hər hansı münasib üsulun seçilməsi və tətbiqi istifadəçilərin konkret tələblərindən asılı olur.

Ümumiyyətlə, axtarışın effektivliyinin artırılması üsullarını dolğunluq və dəqiqlik göstəricilərinin yaxşılaşdırılması kimi iki sinfə bölmək olar.

5.2. Axtarışın dolğunluğunun yaxşılaşdırılması üsulları

İstifadəçi öz sorğusuna hər hansı dərəcədə aidiyyəti olan bütün mümkün sənədlər çoxluğunu əldə etmək istədikdə axtarışın dolğunluğunun yaxşılaşdırılması üsulları tətbiq olunur. Məsələn, ekspert tərəfindən onu maraqlandıran sahəyə aidiyyəti olan bütün patentlərin axtarılması sorğusuna müfəssəl cavabın alınması arzusu buna nümunə ola bilər.

Axtarışın dolğunluq göstəricisinin yaxşılaşdırılması üçün tətbiq olunan üsullar istifadəçinin sorğusunun açar sözləri ilə informasiya massivində olan sənədlərin terminlərinin əlavə üst-üstə düşməsinə təmin etməyə imkan verir. Bu məqsədlə sorğuların və sənədlərin açar sözləri və terminləri ya başqa terminlərlə əvəz olunur, ya da onlara yeniləri əlavə olunur. Başqa sözlə, sorğuların və sənədlərin terminlər çoxluğu genişləndirilir.

Terminlər çoxluğunun genişləndirilməsi üçün bir neçə üsul tətbiq olunur:

- sinonimlər və tezaurus lüğətlərinin istifadəsi;
- assosiativ sözlər çoxluğunun istifadəsi;
- ehtimala əsaslanan indeksləşdirmə;
- bibliografik məlumatların istifadəsi.

Yuxarıda sadalanan üsullardan daha məşhuru *sinonimlər və tezaurus lüğətlərinin istifadə olunması* üsuludur. Bu lüğətlər sinonim və ya ekvivalent sözlərin sinifləri çoxluğundan ibarət olur. Hər sinifdə bir-biri ilə sinonim və ya tezaurus təşkil edən terminlər, eyni bir obyektin və ya hərəkətin təsvirini verən, mənaca bir-birinə çox yaxın olan, eyni mövzu sahəsinə aid edilən sözlər qruplaşdırılır.

Belə lüğətlərin istifadəsi axtarış prosesində sorğunun hər hansı termininin müvafiq sinonimlər sinfinin identifikatoruna dəyişdirilməsinə imkan verir. Nəticədə verilmiş hər hansı bir termin bu sinfə daxil olan terminlər çoxluğu ilə əvəz edilmiş olur.

Terminlərin əlavə üst-üstə düşmələrini təmin etmək üçün tətbiq olunan növbəti üsul *assosiativ sözlərin istifadəsi* üsuludur. Bu üsula əsasən sorğuda göstərilən və sənədi təsvir edən hər bir termin üçün onunla assosiasiya təşkil edən əlavə terminlər çoxluğu əlavə olunur.

Terminlərin assosiativlik göstəricisinin müəyyən edilməsi üçün indeksləşdirmə üsullarından, məsələn, statistik üsuldən istifadə etmək olar. Belə üsullara əsasən terminlər çoxluğu üçün $A = \{a_{ij}\}_{m \times n}$ assosiativlik matrisi tərtib olunur. Bu matrisin sətir və sütunlarının kəsişməsində yerləşən a_{ij} elementi i və j terminləri cütü üçün assosiativlik dərəcəsinin (assosiativlik əmsalının) qiymətini müəyyən edir.

İki i və j terminlərinin assosiativlik əmsalı bu terminlərin informasiya massivinin bütün sənədləri üzrə rastgəlinmə tezliklərinin hasillərinin cəmi şəklində hesablanır:

$$a_{ij} = \sum_{k=1}^n f_{ik} \cdot f_{jk} , \quad (8.1)$$

burada a_{ij} – i və j terminlərinin assosiativlik əmsalı, f_{ik} – i termininin k sənədində rastgəlinmə tezliyi, n – massivdə olan sənədlərin sayıdır.

Yuxarıda qeyd olunduğu kimi, i və j terminlərinin k sənədində rastgəlinmə tezlikləri f_{ik} və f_{jk} $[0,1]$ intervallında qiymətlər alır. Lakin (8.1) düsturuna uyğun olaraq, hesablama nəticəsində a_{ij} assosiativlik əmsalının qiyməti çox böyük ola bilər. Ona görə də bu əmsalın qiymətini normallaşdırmaq məqsədilə aşağıdakı düstur tətbiq olunur:

$$a_{ij} = \frac{\sum_{k=1}^n f_{ik} \cdot f_{jk}}{\sum_{k=1}^n (f_{ik})^2 + \sum_{k=1}^n (f_{jk})^2 - \sum_{k=1}^n f_{ik} \cdot f_{jk}} \quad (8.2)$$

Terminlərin assosiativlik göstəricilərinin qiymətləri üçün δ sərhəd qiyməti müəyyən olunur. Əgər $a_{ij} \geq \delta$ olarsa, onda i və j terminləri assosiativ terminlər kimi qəbul olunur.

Ehtimala əsaslanan indeksləşdirmə zamanı əvvəlcə terminlərin sənədlərdə mövcud olub-olmaması müəyyənləşdirilir, bundan sonra terminlərlə sənədlər arasındakı münasibətlərin əsasında bu terminlərə uyğun olan tematik siniflərin identifikatorları həmin sənədlərə mənimsədilir, yəni sənədlər bu terminləri özündə saxlayan tematik siniflərə aid edilir.

Nəhayət, *bibliografik məlumatların istifadə olunması* üsulları açar sözlərin sayını artırmağa, bunun nəticəsində isə axtarış prosesində əhatə olunan sənədlər dairəsini genişləndirməyə imkan verir. Bibliografik məlumatlar qismində müəlliflərin soyadı, nəşriyyatların adı, sənədlərdə

rast gəlinən istinadlar, sitatlar və s. kimi rekvizitlər istifadə olunur. Onlar açar sözlər qismində sənədlərə (onların indekslərinə) əlavə edilərək yazılır.

5.3. Axtarışın dəqiqliyinin yaxşılaşdırılması üsulları

Axtarışın dəqiqliyinin yaxşılaşdırılması üsulları sorğuya daha az uyğun olan, eləcə də təsadüfi və lazımsız sənədləri istisna etmək və daha relevant sənədləri saxlamaq yolu ilə sistem tərəfindən tapılaraq istifadəçiyə verilən sənədlər çoxluğunu (siyahını) daraltmağa xidmət edir.

Yuxarıda qeyd olunduğu kimi, axtarışın dolğunluğunun yaxşılaşdırılması sorğuların və sənədlərin təsviri üçün istifadə olunan terminlər çoxluğunun genişləndirilməsi, habelə onların digər qohum terminlərlə əvəz olunması yolu ilə əldə edilir. Lakin dolğunluqdan fərqli olaraq, axtarışın dəqiqliyi daha dar mənalı terminlərin və ya bir neçə terminin (yəni söz birləşmələrinin) birgə tətbiq olunması yolu ilə yaxşılaşdırıla bilər. Qeyd olunmalıdır ki, dolğunluğun yaxşılaşdırılması üçün əlavə olunan terminlər sorğuya «OR» məntiqi əməli ilə, dəqiqliyin yüksəldilməsi üçün isə «AND» məntiqi əməli ilə daxil edirlər.

Dar mənalı terminlərin və söz birləşmələrinin seçilməsi üçün xüsusi üsullardan istifadə olunur. Lakin bu üsullar bir sıra üstün və çatışmayan cəhətlərə malikdirlər.

Söz və ya termin birləşmələrinin yaradılması üçün statistik üsulun məğzi aşağıdakıdan ibarətdir. Fərz edilir ki, məsivdəki sənədlərdə söz birləşmələrinin, yəni terminlərin birgə rast gəlinməsi tezliyi bu terminlərin ayrı-ayrılıqda rast gəlinməsi tezliyindən yüksək olmur. Onda iki termin arasındakı bağlılıq dərəcəsini aşağıdakı kimi müəyyən etmək olar:

$$C^{ij} = \frac{F^{ij}}{F^i \cdot F^j} \cdot N, \quad (8.3)$$

burada F^{ij} – massivin sənədlərində i və j terminləri cütlüyünün birgə rast gəlinməsi tezliyi, F^i və F^j isə massivin sənədlərində uyğun olaraq i və j terminlərinin ayrı-ayrılıqda rast gəlinmə tezlikləri, N – massivin sənədlərində olan terminlərin və ya sözlərin ümumi sayıdır.

Massivin sənədlərinin mətnlərindən müvafiq söz birləşmələrinin seçilməsinin effektivliyini yüksəltmək üçün əvvəlcə sənədin sözləri çoxluğundan bütün xidməti sözlər çıxarılır, sonra isə qalan sözlər arasından birgə rast gəlinmə tezliyi və əlaqəlik əmsalı yüksək olan terminlərin kombinasiyaları seçilir. Bu zaman tezliklər üçün müəyyən sərhəd qiymətləri daxil edilir. F^{ij} birgə rast gəlinmə tezliyi və C^{ij} əlaqəlik dərəcəsi verilmiş hər hansı sərhəd qiymətlərindən yuxarı olan bütün terminlər söz birləşmələri kimi qəbul edilir.

Söz birləşmələrinin seçilməsi üçün statistik üsullar heç də həmişə yararlı olmur. Belə ki, onlar sənədin məzmununun həddən artıq daralmasına və dolğunluğun itməsinə gətirib çıxara bilər.

Sintaktik və semantik üsullar söz birləşmələrini təşkil edən komponentlərin qrammatik və mətnin strukturlarının isə linqvistik təhlilinə əsaslanırlar. Bu üsullar təbii dillərlə bağlı olur. Belə ki, mətnlərin sintaktik və semantik xassələri onların hazırlandığı dillərin xüsusiyyətlərindən bilavasitə asılı olur. Sintaktik və semantik üsulların az inkişaf etməsinə baxmayaraq, hal-hazırda bir sıra mövcud tam-mətnli axtarış sistemlərində bu üsullardan istifadə olunmağa başlanmışdır.

5.4. Relevantlığın yüksəldilməsi üçün istifadəçi ilə əks əlaqənin istifadəsi

Adətən, axtarış sistemlərində sənədlər terminlər vasitəsilə qeyri-dəqiq təqdim olunurlar. Ona görə də istifadəçilər axtarış prosesini qeyri-dəqiq, yəni natamam sorğular əsasında aparmağa məcbur olurlar. Bu isə təbii ki, arzu olunan nəticələri əldə etməyə imkan vermir. Amma sonra prosesin növbəti mərhələlərində istifadəçilər iterativ yolla öz sorğularını dəqiqləşdirərək tədricən effektivliyi artırma və axtarışın nəticəsini yaxşılaşdırmağa bilirlər.

Bunun üçün axtarış sistemlərində istifadəçi ilə əks əlaqə mexanizmi reallaşdırılır. Bu mexanizm istifadəçiyə onun ilkin sorğusuna əsasən sistem tərəfindən tapılmış sənədlərin relevantlıq dərəcəsini qiymətləndirərək öz fikrini sistemə ötürməyə imkan verir. Bu, iki və ya çoxsəviyyəli münasibətlərin əsasında həyata keçirilir.

İkisəviyyəli münasibətlərdə istifadəçi aşağıdakı iki variantdan birini göstərə bilər: sistem tərəfindən verilmiş sənədlər sorğuya ya relevant olur, ya da qeyri-relevant. Çoxsəviyyəli münasibətlər istifadə olunduqda isə bu iki variantdan əlavə həm də sənədlərin qismən relevant olması münasibəti müəyyən edilə bilər. Əks əlaqə üsullarının reallaşdırılmasının sadələşdirilməsi üçün, adətən, ikisəviyyəli relevantlıq münasibətlərindən istifadə olunur.

Hazırda iki əks əlaqə üsulu istifadə olunur:

- sorğunun təqdimatının dəyişdirilməsi;
- sənədin təqdimatının dəyişdirilməsi.

Sorğunun təqdimatının dəyişdirilməsi üsulları cari seansda axtarışın effektivliyini artırmaq üçün nəzərdə tutulmuşdur, onlar digər sorğular üzrə axtarışın effektivliyinə təsir etmir.

Sənədin təqdimatının dəyişdirilməsi üsulları isə sənədlərin təqdimat formasını dəyişdirir və ona görə də yalnız cari axtarış seansda deyil, həmçinin sonrakı seanslarda da axtarışın nəticəsini yaxşılaşdırmağa imkan verir.

Sorğunun təqdimatının dəyişdirilməsi üsullarının köməyi ilə axtarışın effektivliyinin artırılması üçün üç üsuldan istifadə olunur:

- terminlərin çəkilərinin dəyişdirilməsi;
- sorğunun genişləndirilməsi;
- sorğunun parçalanması (hissələrə ayrılması).

Terminlərin çəkilərinin dəyişdirilməsi üsulu sorğuya daxil edilmiş terminlərin çəkilərinin yaxşılaşdırılmasını həyata keçirir. Terminlərin çəkilərinin korrektə edilməsi üçün sorğu vektoru ilə müsbət əks əlaqəli sənədlərin vektorlarını toplamaq lazımdır. Terminlərin çəki əmsallarını, həmçinin sorğu vektorundan mənfi əks əlaqəli sənədlərin vektorlarının çıxılması yolu ilə də korrektə etmək olar.

Qeyd olunmalıdır ki, *müsbət əks əlaqəli sənədlər* dedikdə axtarış nəticəsində istifadəçiyə verilmiş və istifadəçi tərəfindən sorğuya relevant sənəd kimi qiymətləndirilmiş sənədlər başa düşülür.

Axtarış sisteminin verdiyi sənədlər içərisindən istifadəçi tərəfindən qeyri-relevant sənəd kimi qeyd olunmuş sənədlər isə *mənfi əks əlaqəli sənədlər* kimi qəbul olunur.

Sorğu korrektə edildikdən sonra axtarış sistemi tərəfindən verilmiş relevant sənədlərin sayı korreksiyadan əvvəlkindən çox olmalıdır. Belə ki, əvvəlki relevant sənədlərdən əlavə əks əlaqənin köməyi ilə onlar arasından istifadəçi tərəfindən müsbət qeyd olunmuş relevant sənədlərə oxşar yeni sənədlər də bu çoxluğa əlavə olunur.

Relevant sənədlərin keyfiyyətində və sayında arzu olunan səviyyə əldə edilənə qədər istifadəçi bu prosesi davam etdirə bilər.

Sorğunun genişləndirilməsi üsulu ilkin sorğuya yeni terminlər əlavə etmək yolu ilə onu daha əhatəli edir. Belə ki, əvvəlcə müsbət əks əlaqəli sənədləri təsvir edən terminlər çoxluğu seçilir. Bu çoxluq terminlərin vaciblik dərəcələrinə görə nizamlandıqdan sonra siyahının əvvəlindən verilmiş sayda termin götürülür və sorğuya əlavə edilir.

Sorğunun parçalanması (hissələrə ayrılması) üsulu axtarış nəticəsində tapılmış sənədlər arasında müsbət əks əlaqəli müxtəlif xarakterli sənədlər olduqda tətbiq olunur. Belə hallarda istifadəçiyə verilmiş sənədlər çoxluğu eyni xarakterə malik sənədlərdən ibarət bir neçə alt çoxluğa (klasterə) bölünür. Əgər bu mümkün olarsa, onda istifadəçinin sorğusu elə alt sorğulara (yəni ilkin sorğunun alt hissələri olan bir neçə ayrı-ayrı sorğulara) bölünür ki, hər bir alt sorğu ayrı-ayrı sənədlər alt çoxluğuna uyğun olsun.

Bundan sonra hər bir alt sorğu üzrə ayrı-ayrılıqda axtarış aparılır, onların terminlərinin çəki əmsalları yaxşılaşdırılır və ya bu alt sorğular əvvəl qeyd olunmuş üsulların köməyi ilə genişləndirilir. Beləliklə, axtarışın effektivliyinin səviyyəsi yüksəldilmiş olur.

Sənədlərin təqdimatının dəyişdirilməsi üsulları istifadəçi ilə əks əlaqənin köməyi ilə axtarış sisteminin indekslər bazasında sənədlərin təqdim olunması vektorunu yaxşılaşdırmağa imkan verir. Bu zaman terminlərin çəki əmsalları elə şəkildə korrektə edilir ki, tapılmış relevant sənədlərin vektoru sorğunun vektoruna müəyyən qədər yaxınlaşmış olsun, qeyri-relevant sənədlərin vektorları isə əksinə ondan uzaqlaşsın.

Sənədlərin təqdim olunması vektorlarının belə korreksiyyasından sonra növbəti dəfə analogi axtarış sorğusu daxil edildikdə sistem əvvəlki axtarış seanslarına nisbətən daha relevant sənədləri tapıb istifadəçiyə verir.

VI FƏSİL

İNFORMASIYA AXTARIŞININ MODELLƏRİ

**İnformasiya axtarışının modelləşdirilməsinə
yanaşmalar**

Nəzəri-çoxluqlar modeli

Vektorlar modeli

Statistik tezliklər modeli

Ehtimallar modeli

Linqvistik model

Semantik model

Qeyri-səlis model

6.1. İnformasiya axtarışının modelləşdirilməsinə yanaşmalar

Mütəxəssislərin fikrinə görə, internetdə fəaliyyət göstərən informasiya-axtarış sistemlərinin böyük əksəriyyəti hər hansı riyazi modelə əsaslanmır. Onlarda yalnız istifadəçinin daxil etdiyi sorğuya uyğun olan müəyyən sənədlərin tapılıb verilməsini təmin edən alqoritmlər reallaşdırılmış olur. Lakin axtarışın keyfiyyətinin yaxşılaşdırılması, məlumatın həcmnin böyüklüyü, istifadəçi sorğularının axınının intensivliyi ilə bağlı məsələlərin həlli zamanı təcrübəli alqoritmlərlə yanaşı nəzəri riyazi aparata əsaslanan axtarış modelinə (sadə olsa belə) zərurət yaranır.

Axtarışın modeli dedikdə əsasında hansı sənədlərin tapılması və nizamlanması barədə qərar qəbul etməyə imkan verən zəruri qaydaların yaradıldığı real vəziyyətin müəyyən sadələşdirilməsi başa düşülür. Model qəbul edildikdən sonra məsələnin həlli üçün zəruri olan parametrlər həqiqi mənalərini ifadə edir və başa düşülən olur.

İnformasiya axtarışı modelləri əsas dörd parametrlə xarakterizə olunur:

- sənədlərin və sorğuların təqdim edilməsi;
- sənədin istifadəçinin sorğusuna relevantlığının qiymətləndirilməsi üçün tətbiq olunan qarşılıqlı müqayisə üsulları;
- axtarışın nəticələrinin nizamlanması üsulları;
- relevantlığın istifadəçi tərəfindən qiymətləndirilməsini təmin edən əks əlaqə mexanizmləri.

Sənədlərin təqdim olunması, indeksləşdirilməsi, axtarış obrazlarının yaradılması, eləcə də reallaşdırılan informasiya

axtarışı üsullarından və vasitələrindən asılı olaraq, informasiya axtarışının bir sıra modellərini fərqləndirirlər. Bütövlükdə bu modellərin effektivliyi qeyd olunan üsul və vasitələrin reallaşdırılmasının keyfiyyəti ilə müəyyən olunur.

Sənədli informasiya axtarışının ümumiləşdirilmiş modelini aşağıdakı kortej şəklində vermək olar:

$$\{D, Q, F, R(d, q)\}.$$

Burada D – sənədlərin təqdimatları çoxluğu, Q – informasiya sorğularının təqdimatları çoxluğu, F – sənədlərin və sorğuların təqdimatlarının təhlili, qiymətləndirilməsi və müqayisəsi, onlar arasındakı münasibətlərin modelləşdirilməsi vasitələri, $R(d, q)$ – nizamlama funksiyasıdır. $R(d, q)$ funksiyası hər bir $d \in D$ sənədinə və $q \in Q$ sorğusuna qarşı həqiqi ədədlər qoyur, $q \in Q$ sorğusuna görə $d \in D$ sənədinin axtarış nəticəsində verilmiş sənədlər çoxluğunda sırasını müəyyən edir.

Aşağıda informasiya axtarışının daha məşhur olan modellərinə baxılır:

- nəzəri-çoxluqlar modeli;
- vektorlar modeli;
- statistik tezliklər üsulu;
- ehtimallar modeli;
- linqvistik model;
- semantik model;
- qeyri-səlis model.

6.2. Nəzəri-çoxluqlar modeli

Nəzəri çoxluqlar (bul) modelləri tammətli axtarışı reallaşdıran ən ilk və çox istifadə olunan modellərdir. Bu

modellərə əsasən, axtarış sorğusundakı söz sənəddə rast gəlinirsə, onda həmin sənədin axtarılan sənəd olduğu qəbul edilir. Əks halda sənəd sorğuya relevant hesab olunur.

Nəzəri-çoxluqlar modelində sənədlərin və sorğuların təsvir edilməsi üçün bul dəyişənlərindən istifadə olunur. Məhz bu səbəbdən çox vaxt nəzəri-çoxluqlar modellərini “*bul modelləri*” adlandırırlar. Beləliklə, bul modellərində sənədlər onun indeksində olan və bul dəyişənləri (məntiqi dəyişənlər) kimi qəbul edilən terminlər çoxluğunun köməyi ilə təqdim olunurlar. Əgər hər hansı termin sənəddə rast gəlinirsə, onda bu termin üçün uyğun məntiqi dəyişən “True”, əks halda “False” qiymətini alır. Terminlərə hər hansı çəki əmsallarının mənimsədilməsinə icazə verilmir. Belə ki, burada yalnız terminin sənədə daxil olub-olmaması (True və ya False) faktı təsbit olunur.

Bu modelə uyğun olaraq, istifadəçilər öz sorğularını məntiqi (bul) ifadələr şəklində formalaşdırırlar. Sorğulara daxil olan terminlər AND, OR, NOT və s. kimi məntiqi əməllərin köməyi ilə bir-birinə bağlanırlar. Axtarış sistemi tərəfindən istifadəçinin sorğusu qəbul edildikdən sonra informasiya fəzasının hər bir sənədi üçün seçim statusunun qiymətləri – retrieval status values (RSV dəyişəni) hesablanır. RSV dəyişəninin qiymətləri informasiya fəzasının sənədlərinin istifadəçinin sorğusuna uyğunluğu dərəcələrini müəyyən etməyə imkan verir. RSV dəyişəni “1” və ya “0” qiymətlərindən birini alır. Əgər verilmiş sorğu üçün məntiqi ifadənin qiyməti “True” olarsa, onda RSV dəyişəni “1” qiymətini alır və müvafiq sənədlər verilmiş sorğuya relevant kimi qəbul edilir. Əgər ifadənin qiyməti “False” olarsa, onda $RSV=0$ qiymətini alır və müvafiq sənədlər sorğuya qeyri-relevant hesab olunur.

AND operatoru sorğuya cavab olaraq verilən sənədlərin sayını çox kəskin sürətdə azalda bilər. Bu zaman hər şey axtarış terminlərinin indekslər bazası üçün nə qədər xarakterik olmasından çox asılı olur. OR operatoru isə əksinə sorğunun özünü doğrultmayan genişlənməsinə gətirib çıxara bilər. Belə olduqda isə faydalı informasiyanın lazımsız məlumatlar (informasiya səs-küyü) arasında itib batmasına səbəb ola bilər.

Bu modeli əsasında qurulmuş informasiya-axtarış dilinin müvəffəqiyyətlə tətbiq edilməsi üçün sistemin leksikasını və tematik istiqamətini yaxşı bilmək tələb olunur. Adətən, belə İAD olan sistemlər üçün xüsusi mürəkkəb lüğətlərə malik sənədli leksik məlumat (indekslər) bazaları yaradılır. Tezaurus adlanan bu lüğətlər terminlər arasındakı əlaqələr haqqında məlumatı özündə saxlayır.

Qeyd etmək lazımdır ki, məntiqi modellər reallaşdırma üçün çox sadədir və bu baxımdan hazırda onlar bir çox kommersiya əsaslı (özəl) axtarış sistemlərində tətbiq olunurlar. Onlar məntiqi əməllərin köməyi ilə istifadəçilərə öz sorğularını ixtiyari mürəkkəb ifadələr şəklində formalaşdırmağa imkan verir.

Həmçinin qeyd olunmalıdır ki, məntiqi modeldə axtarışın effektivliyi, adətən, yüksək olmur. Belə ki, terminlərə çəki əmsallarının mənimsədilməməsi və relevant sənədlər qismində tapılmış bütün sənədlər üçün RSV dəyişəninin eyni qiymətlər alması səbəbindən axtarışın nəticəsini nizamlamaq və sıraya düzmək mümkün olmur. Ona görə də, tapılıb istifadəçiyə verilmiş sənədlərin siyahısında (onların sayı minlərlə, bəzən isə yüz minlərlə ola bilər) daha vacib sənədlər siyahının ortasına və ya axırına düşə və istifadəçinin diqqətindən kənar qala bilərlər.

Bu modeldə axtarışın effektivliyini artırmaq məqsədilə çox vaxt istifadəçi ilə əks əlaqə üsulundan istifadə olunur.

İAS tapılmış sənədlərin siyahısını istifadəçilərə verərkən sistem onlardan siyahının əvvəlində yerləşən bir neçə sənədin həqiqətən relevant olub-olmamasını göstərməyi xahiş edir.

Məntiqi modelin dəyişdirilmiş (modifikasiya olunmuş) variantı çəkiyə (qeyri-səlis çoxluqlar və münasibətlər nəzəriyyəsinin köməyi ilə qiymətləndirilir) əsaslanan bul axtarışıdır. Belə axtarışın ideyası olduqca sadədir. Fərz edilir ki, termin sənədin məzmununu hər hansı dəqiqliklə təsvir edir və bu dəqiqlik terminin həmin sənəd üçün çəkisi kimi qəbul olunur. Bu zaman həm sənədin terminlərinin, həm də sorğuya daxil olan terminlərin çəkisini müəyyənləşdirmək olar. Sorğular yuxarıda təsvir olunmuş informasiya-axtarış dilinin köməyi ilə formalaşdırıla, axtarışın nəticəsi isə istifadəçiyə verilən zamanı sənədlərin sorğuya yaxınlıq dərəcəsinə uyğun olaraq nizamlana bilər. Burada yaxınlığın ölçülməsi elə aparılır ki, məntiqi axtarış modeli çəkiyə əsaslanan bul axtarışının xüsusi halını təşkil etsin.

6.3. Vektorlar modeli

Vektorlar modeli informasiya axtarışının atası hesab olunan Gerard Salton tərəfindən 1968-ci ildə SMART axtarış sistemində reallaşdırılmışdır. Bu model çərçivəsində nəticənin nizamlanması təbii statistik müşahidəyə əsaslanmışdır.

Vektorlar modelində (onu bəzən *cəbri model* adlandırırlar) sənədlər informasiya fəzasının vektorları çoxluğu ilə təqdim olunurlar. n normallaşdırılmış vektordan ibarət bazislə müəyyən olunan bu fəzada hər bir sənəd bir n -ölçülü vektorla təqdim edilir. Bu vektorun birinci elementi vektorlar fəzasında ölçməyə uyğun olaraq birinci terminin sənəddəki çəkisini, ikinci elementi ikinci terminin çəkisini və s. əks etdirir.

Vektorlar modelində qəbul edilmişdir ki, fəzanı əmələ gətirən terminlərin vektoru ilə sənədlərin vektorları ortoqonal olmalıdır. Burada terminlər arasında mövcud olan qarşılıqlı əlaqələr nəzərə alınmır. Adətən, informasiya axtarışının vektorlar modelində bir çox anlayışlardan istifadə olunur.

İnformasiya axını və ya massivi $n*m$ ölçülü matris şəklində təqdim olunur. Burada n – sənədlərin, m isə massivdə olan sənədlərin təsvir olunması üçün istifadə olunan terminlərin ümumi sayıdır. Matrisin sətirləri qismində sənədlərin axtarış obrazları çıxış edir: $D = \{d_1, d_2, \dots, d_N\}$.

İnformasiya massivinin ümumi terminləri çoxluğu *lüğət* adlanır və $T = \{t_1, t_2, \dots, t_m\}$ ilə işarə olunur.

Sənədlər m ölçülü vektorlar şəklində təqdim edilirlər. Əgər hər hansı termin verilmiş sənədə daxildirsə, onda vektorun müvafiq müvafiq elementinə 1, əks halda isə 0 qiyməti mənimsədilir. Adətən, xətti indeksləşdirmə modelində bütün axtarış əməliyyatları sənədlərin vektorlar şəklində təqdim edilmiş axtarış obrazları üzərində yerinə yetirilir, lakin bu zaman sadəlik naminə bu obrazları sənəd adlandırırlar.

İstifadəçi sorğusu da l ölçülü vektorla təqdim olunur: $Q = \{q_1, q_2, \dots, q_l\}$. L – sorğuya daxil edilmiş açar sözlərin sayıdır.

Axtarış prosesində sənədin sorğuya uyğunluğunu müəyyən etmək üçün sənəd və sorğu vektorlarının skalyar hasili hesablanır:

$$R = Q \cdot D, \quad (6.1)$$

burada Q – sorğu vektoru, R – sistemin sorğuya cavabıdır. Alınan nəticə nə qədər böyük olarsa, sənəd bir o qədər sorğuya relevant hesab edilir.

Nəticənin yaxşılaşdırılması məqsədilə relevantlığa görə sorğunun korrektə edilməsi mexanizmindən istifadə olu-

nur. Bu mexanizmin mahiyyəti ondan ibarətdir ki, axtarış prosedurası interaktiv və iterativ xarakter daşıyır. Belə ki, ilkin axtarış yerinə yetirildikdən sonra sistem tərəfindən təqdim olunmuş sənədlər arasından istifadəçinin relevant hesab etdiyi sənədləri göstərməsi ondan xahiş (təklif) olunur. İterasiyanın sonrakı mərhələsində sistem tərəfindən qeyd olunmuş sənədlərə daxil olan terminlər vasitəsilə istifadəçinin sorğusu genişləndirilir (və ya dəqiqləşdirilir) və yenidən axtarış aparılır. Axtarış prosedurası bu qayda ilə o vaxta qədər davam etdirilə bilər ki, istifadəçinin fikrinə görə artıq bundan sonra axtarışın nəticəsini yaxşılaşdırmağa ehtiyac olmasın. Relevantlığa görə sorğunun korrektə edilməsi üsulu sorğuların dəqiqləşdirilməsi məqsədilə kifayət qədər geniş tətbiq olunur. Bəzi sistemlərdə istifadəçi belə bir proseduranın tətbiq olunduğunu hiss etmir.

Vektorlar modelinin üstünlüyü, əsasən, onun reallaşdırılmasının asan və sadə olmasından ibarətdir. Bu model həmçinin tapılıb verilmiş sənədlərin relevantlığının qiymətləndirilməsi üçün istifadəçi ilə əks əlaqəni asan reallaşdırmağa imkan verir.

6.4. Statistik tezliklər üsulu

Tutaq ki, $D = \{d_1, d_2, \dots, d_n\}$ – informasiya massivində olan sənədlər çoxluğu, n – informasiya massivində olan sənədlərin sayı, $T = \{t_1, t_2, \dots, t_m\}$ – sənədlərin məzmunlarını daha dolğun təsvir etməyə və ya onların hansı tematikaya mənsub olmasını müəyyən etməyə imkan verən terminlər çoxluğu, $W = \{w_{ij}\}_{n \times m}$ – terminlər və sənədlər arasındakı münasibətləri müəyyən edən matrisdir. Bu matrisin w_{ij} ele-

mentinin qiyməti informasiya massivinin bütün sənədlərinin xarakteristikalarını nəzərə almaqla t_j termininin d_i sənədində çəkisini müəyyən edir. Burada w_{ij} , $i = \overline{1, n}$, $j = \overline{1, m}$ elementləri $[0,1]$ intervalında qiymətlər alır, $w_{ij} = 0$ qiyməti t_j termininin d_i sənədində ümumiyyətlə rast gəlinmədiyini, $w_{ij} = 1$ qiyməti isə t_j termininin d_i sənədinin məzmununa tamamilə (100%) uyğun olduğunu göstərir.

Terminlərin w_{ij} çəki əmsalları sənədlərin istifadəçilər tərəfindən açar sözlər çoxluğu şəklində tərtib olunmuş sorğulara relevantlıq dərəcəsinin müəyyənləşdirilməsi üçün istifadə olunur. Yuxarıda qeyd olunduğu kimi, informasiya massivinin sənədləri axtarış sisteminin bazasında T terminlər çoxluğunun alt çoxluğu ilə təsvir olunur. Ona görə də bu modelə uyğun olaraq, axtarış prosesində elə sənədlər tapılıb istifadəçilərə verilməlidir ki, onların sorğusunda göstərilmiş açar sözlər bu sənədlərə termin qismində ən böyük çəki əmsalları ilə daxil olsunlar.

İndi isə w_{ij} çəki əmsallarının qiymətlərinin müəyyən edilməsi üçün statistik üsula baxaq. Bu üsul iki mərhələdən ibarətdir. Birinci mərhələdə hər bir terminin ayrı-ayrı sənədlərə daxilolma tezlikləri hesablanır.

Tutaq ki, $f_{ij} - t_j$ termininin d_i sənədinə daxilolma tezliyidir. Onda:

$$f_{ij} = \frac{m_{ij}^t}{m_i^w}, \quad (6.2)$$

burada $m_{ij}^t - t_j$ termininin d_i - sənədinə daxil olmalarının sayı, $m_i^w - d_i$ sənədində olan sözlərin ümumi sayıdır. Bu-

radan görünür ki, sənəddə daha çox rast gəlinən istənilən söz bu sənədin məzmununu ən dəqiq əks etdirən vacib termin kimi qəbul oluna bilər.

Lakin praktika göstərir ki, sənəddə ən böyük rastgəlinmə tezliyinə malik olan sözlər heç də həmişə bu sənəd üçün vacib termin olmurlar. Belə sözlərə nümunə kimi xidməti sözləri, bağlayıcıları, əvəzlilikləri və s. göstərmək olar.

Sənədlərin məzmunlarının Zipf qanuna uyğun olaraq təhlilinin nəticələrinə əsasən demək olar ki, əgər sənəddə hər hansı söz böyük rastgəlinmə tezliyinə malikdirsə, lakin termin deyilsə, onda bu söz massivin digər bir çox sənədlərində də böyük tezliklə rast gəlinməlidir. Məhz bu xassə sənədlər üçün vacib terminlərin seçilməsi prosesində istifadə oluna bilər. Ona görə də ikinci mərhələdə informasiya massivinin bütün sənədləri təhlil olunur və terminlərin massivin digər sənədlərində rastgəlinmə xarakteristikaları nəzərə alınmaqla onların həmin sənəd üçün çəki əmsalı müəyyən edilir.

Tutaq ki, $n_j^d - t_j$ termininin rast gəlinmədiyi sənədlərin sayıdır. n ilə n_j^d arasındakı asılılığı təhlil etmək yolu ilə t_j termininin verilmiş sənəd üçün hansı dərəcədə əhəmiyyətli olduğunu müəyyən etmək olar. Burada n_j^d kəmiyyətini terminlərin sənəd tezliyi adlandırırlar.

Qeyd olunmalıdır ki, n_j^d kəmiyyətinin qiyməti nə qədər kiçik olarsa, terminin sənəddəki çəkisi bir o qədər böyük

olur. $\log \frac{n}{n_j^d}$ kəmiyyəti t_j termininin sənədləri diskrimi-

nasiya edib-etməməsi amilini özündə əks etdirən yaxşı indeksator rolunu oynayır, yəni bu terminin əhəmiyyət kəsb etdiyi sənədləri onun rast gəlinmədiyi digər sənədlərdən fərqləndirməyə imkan verir.

Bu kəmiyyət əks sənəd tezliyi adlanır və aşağıdakı kimi işarə olunur:

$$f_j^d = \log \frac{n}{n_j} \quad (6.3)$$

Terminin sənəddəki tezliyi və əks sənəd tezliyi kəmiyyətlərini tezlik üzrə vahid indeksləşdirmə modeli çərçivəsində birləşdirmək olar:

$$w_{ij} = f_{ij} \cdot f_j^d \quad (6.4)$$

(6.2) və (6.3) ifadələrini (6.4) düsturunda yerinə qoysaq, informasiya fəzasının bütün digər sənədlərinin xarakteristikalarını nəzərə almaqla t_j termininin d_i sənədindəki w_{ij} çəki əmsalının hesablanması üçün düsturu almış olarıq:

$$w_{ij} = \frac{m_{ij}^t}{m_i^w} \cdot \log \frac{n}{n_j} \quad (6.5)$$

(6.5) düsturundan görünür ki, t_j termininin d_i sənədində rast gəlinmə tezliyi nə qədər böyük, eləcə də bu terminin rast gəlinədiyi sənədlərin sayı nə qədər az olarsa, onda t_j termini d_i sənədində bir o qədər böyük çəkiyə malik olar. Başqa sözlə, əgər d_i sənədi t_j termininin cəmləşdiyi sənəd olarsa, onda t_j termini bu sənəd üçün əhəmiyyət kəsb edən termin rolunu oynayır.

Qeyd olunmalıdır ki, terminlərin sənədlərə daxilolma tezlikləri əsasında indeksləşdirmə üsulu, əsasən, axtarış prosesinin bir xarakteristikasını – axtarışın dolğunluğunu yüksək səviyyədə təmin etməyə imkan verir. Lakin məsivinin ayrı-ayrı sənədlərində terminlərin cəmlənməsi amili eyni zamanda axtarışın yüksək dəqiqliyinin əldə olunması üçün də istifadə oluna bilər.

6.5. Ehtimallar modeli

Ehtimal modelləri 1977-ci ildə reallaşdırılmışdır. Bu modellərə əsasən tapılan sənədin sorğuya relevantlığı həmin sənədin istifadəçi üçün maraqlı olması ehtimalına əsasən qiymətləndirilir.

Ehtimallar modeli axtarış prosesinin nəticəsində tapılan relevant sənədlərin nizamlanması prinsipinə əsaslanır. Belə ki, təqdim olunan sənədlərin istifadəçilər tərəfindən müsbət qiymətləndirilməsi ehtimalına və ya onlar üçün əhəmiyyətinə uyğun olaraq sıraya düzmək olar. Burada aşağıdakı məqamlar nəzərə alınır:

- informasiya massivində olan hər bir sənədin relevantlığı digər sənədlərdən asılı deyildir;
- hər bir sənədin faydalılığı istifadəçi tərəfindən artıq baxılmış relevant sənədlərin sayından asılı ola bilər. Belə ki, baxılmış relevant sənədlərin sayı artdıqca qalan sənədlərin faydalılığı azala bilər.

Ehtimallar modelində terminlər arasında mövcud qarşılıqlı asılılıq və əlaqələr nəzərə alınır, axtarışın əsas parametrləri müəyyənləşdirilir. İlk növbədə bu parametrlərə sorğularda olan terminlərin çəkilərini və "sorğu-sənəd" münasibətlərini aid edirlər.

Bu modeldə hər bir sənədə onun indeksləşdirilməsi üçün istifadə olunan və olunmayan terminləri göstərən binar vektor qarşı qoyulur. Vektor sənədlərin sorğuya relevantlığının və ya qeyri-relevantlığının vəziyyətini əks etdirir. Bu, iki əsas parametrlə müəyyən olunur:

- Prel – sənədin istifadəçinin sorğusuna relevant olması ehtimalı;
- Pnrel – sənədin istifadəçinin sorğusuna qeyri-relevant olması ehtimalı.

Bu parametrlər terminlərin ehtimal olunan çəki əmsallarındır və terminlərin sənəddə faktiki mövcudluğu əsasında hesablanır. Sənədlərin relevantlığı binar xassə olduğundan $P^{rel} = 1 - P^{nrel}$ qəbul edilmişdir.

Beləliklə, nəticədə bu modeldə sənədlər d binar vektoru ilə təsvir olunur. Sənədlərin indeksləşdirilməsi zamanı tematik rubrikatorda olan terminlərin istifadə olunub-olunmaması faktı bu vektor vasitəsilə müəyyən edilir. Onun əsasında isə sənədlərin relevantlıq və ya qeyri-relevantlıq vəziyyətlərini müəyyən edən s_1 və s_2 əmsalları hesablanaraq müvafiq sənədlərə mənimsədilir.

d sənədinin istifadəçinin sorğusuna relevant və ya qeyri-relevant olması ehtimalını Bayes düsturu ilə aşağıdakı kimi ifadə etmək olar:

$$P(s_i / d) = \frac{P(d / s_i) \cdot P(s_i)}{P(d)}, \quad (6.6)$$

burada $P(s_i)$ – sənədin relevantlığının ($i=1$) və ya qeyri-relevantlığının ($i=2$) apriori ehtimalıdır, $P(d / s_i)$ – axtarış nəticəsində verilmiş sənədlər arasından həmin sənədin relevant ($i=1$) və ya qeyri-relevant ($i=2$) olması ehtimalıdır.

Tam ehtimal düsturuna görə:

$$P(d) = P(d / s_1) \cdot P(s_1) + P(d / s_2) \cdot P(s_2). \quad (6.7)$$

$P(d / s_i)$ kəmiyyətinin hesablanması və axtarış funksiyasına daxil edilməsi məqsədilə sorğularda olan terminlərin çəkirlərinin müəyyən olunması üçün aşağıdakı mülahizələr əsasında qurulmuş mexanizmdən istifadə olunur:

- sənədlərin indeksləşdirilməsi zamanı terminlər bir-birindən asılı olmadan istifadə olunur;
- sənədin sorğuya relevantlığı ehtimalı sorğuya daxil edilmiş terminlərin sənədin axtarış obrazlarında olub-olmaması əsasında qiymətləndirilir.

Bu mülahizələrin əsasında aşağıdakı çəki funksiyası daxil edilir:

$$w_i = \log \frac{r_i \cdot (n - n_i - r + r_i)}{(r - r_i) \cdot (n_i - r_i)}, \quad (6.8)$$

burada, n – informasiya massivində olan sənədlərin sayı, $n_i - t_i$ termini rast gəlinmiş sənədlərin ümumi sayı, r – informasiya massivində istifadəçi sorğusuna relevant olan sənədlərin sayı, $r_i - t_i$ termini rast gəlinmiş relevant sənədlərin sayı.

r_i , r və n_i kəmiyyətlərinin qiymətləri çox kiçik və ya sıfır olduqda r_i kəmiyyəti ($r_i + 0.5$) ilə əvəz oluna bilər. r_i və r kəmiyyətlərinin müəyyən edilməsi üçün qabaqcadan bəlli olan və sonradan istifadəçi ilə əks əlaqə yolu ilə dəqiqləşdirilən qiymətlər istifadə olunur.

Bundan əlavə, informasiya massivinin ayrı-ayrı hissələri üzrə alınmış nəticələr də istifadə oluna bilər. İstifadəçi ilə əks əlaqə olmadıqda, sistem tərəfindən avtomatik olaraq relevantlıq ehtimalının ən böyük qiyməti mənimsədilmiş sənədlərin relevant sənədlər kimi qəbul edilməsi və bunun əsasında qiymətləndirmənin aparılması mümkündür.

6.6. Linqvistik model

Linqvistik modeldə leksik indeksləşdirmə bul sorğularının optimallaşdırılması üçün nəzərdə tutulmuşdur. Sifət, isim və ya fel kimi sintaktik siniflərə indeksatorların mənimsədilməsi sistemdə istifadə olunan statistik üsulun keyfiyyətini yüksəldə bilər. Bu halda ifadələrin formalaşdırılması müəyyən edilmiş sintaktik indikatorlara (məsələn, isim – isim və ya sifət – isim şəklində) malik olan cümlələrlə məhdudlaşır.

Sintaktik vahidlərin identifikasiyası üçün sadə sintaktik təhlil prosedurası istifadə oluna bilər. Adətən, ifadələrin elementləri bu sintaktik vahidlər çərçivəsində seçilir.

Leksik indeksləşdirmə daha köhnə və ənənəvi üsul hesab olunur. İstifadə olunduğu dövr ərzində bu üsul açar sözlərə və bütöv ifadələrə görə axtarış sahəsində xeyli təkmilləşdirilmişdir. İnkişaf etmiş leksik axtarış sistemlərində stop-siyahılar, qrammatik sözyaratma və genişlənmiş sorğular dili, eləcə də mətnə sözlərin bir-birinə yaxınlığı və onların mətnin əvvəlinə yaxınlığı parametrlərinin verilməsi imkanları istifadə olunur.

Bir çox sistemlərdə intellektuallıq əl ilə yaradılmış tezaurusların hesabına əldə olunur. Hazırda tezaurusların avtomatik yaradılmasının müxtəlif üsullarının işlənilib hazırlanmasına baxmayaraq, onların generasiya olunduğu xüsusi mühitdən kənarında istifadə üçün bu tezaurusların effektivliyi sual altında qalır. Yəqinliklə demək olar ki, müasir leksik axtarış sistemlərinin imkanları hər hansı dar sahədə informasiya ilə maraqlanan ekspertlərin tələbatını tamamilə ödəyir.

Lakin web mühitdə axtarış sistemlərinin əksər istifadəçiləri üçün leksik axtarışın effektivliyi olduqca qeyri-qənaətbəxş olur. Bunun əsas səbəblərindən biri aşağıdakıdan ibarətdir: ekspertlərdən fərqli olaraq adi istifadəçiyə mövzu sahələrində mövcud açar sözləri istifadə etməklə öz sorğularını formalaşdırmaq çətinidir. Əgər hər hansı mövzu sahəsində sabitləşmiş və qəbul olunmuş terminlər yoxdursa, onda sorğunun formalaşdırılması daha çətin olur. Bundan əlavə, dar mövzu sahəsindən kənara çıxmağa cəhd olunduqda sözlərin sinonimlik və polisemiya (müxtəlif mənalara malik eyni sözlər) xarakteristikaları çox ciddi problem kimi meydana çıxır.

6.7. Semantik model

İnformasiya axtarışının semantik modelinin əsasında süni intellekt sistemlərində biliklərin təqdim edilməsinin semantik modeli durur. İnternetin informasiya fəzasının semantik modeli WWW hörümçək torunun bazasında semantik şəbəkənin reallaşdırılmasını nəzərdə tutur. Məhz buna görə də çox vaxt "semantik model" və "semantik şəbəkə" anlayışlarını eyniləşdirirlər.

Semantik şəbəkə yeni şəbəkə deyil, o, WWW şəbəkəsinin genişlənməsindən və ya tamamlanmasından ibarət olub, kompüterlərə aydın olan, yəni kompüterlər tərəfindən başa düşülən informasiyanı özündə saxlamaq, istifadəçilər üçün kompüterlərlə müvəffəqiyyətlə qarşılıqlı fəaliyyət göstərmək imkanını reallaşdırmalıdır. Semantik şəbəkə məzmunca statik və dinamik, semantik baxımdan nişanlanmış informasiya ehtiyatlarından ibarət olmalıdır. Digər tərəfdən semantik şəbəkə web-texnologiyanın genişlənməsi (inkişafı) olduğundan onun əsasını web-xidmətlər təşkil edir. Ona görə də semantik şəbəkəni çox vaxt semantik web adlandırırlar.

Semantik yanaşma baxımından, kompüterlərdə hazırlanmış informasiyanı iki qrupa ayırırlar:

- insanların başa düşəcəyi və qəbul edə biləcəyi şəkildə olan informasiya;
- kompüterdə emal oluna biləcək informasiya.

Hazırda internet və web-texnologiya bilavasitə insanlar tərəfindən istifadə üçün nəzərdə tutulmuş məlumatların daşıyıcısı kimi formalaşmışdır. Lakin insanların müdaxiləsi olmadan kompüterlər tərəfindən avtomatik emalı mümkün olan informasiyanın toplanması və saxlanması onların ilkin təyinatı olmamışdır.

Semantik modeldə informasiya sistemlərindən (o cümlədən elektron kitabxanalardan) daha dolğun və dəqiq informasiyanın alınmasını təmin etmək məqsədilə kompüter mətnşünaslığı və linqvistikası sahəsində əldə olunmuş nailiyyətlər, qeyri-verbal komponentlərin sənədlərdə rolu, eləcə də mətnlərin avtomatik təhlili sahəsində müasir instrumental vasitələrin genişlənən imkanları nəzərə alınır. Bu modelin köməyi ilə sənədlərdə olan verbal və qeyri-verbal informasiyanın semantik axtarışı texnologiyalarını reallaşdırmağa imkan verir.

Semantik model informasiya texnologiyalarının müasir inkişafı, eləcə də tammətənlı sənədlər kolleksiyalarını saxlayan politematik elektron kitabxanaların yaranması, metatexnologiyaların və internet şəbəkəsinin geniş istifadəsi, semiotik baxımdan müxtəlif növ sənədlərin bütün komponentlərinin və göstəricilərinin məzmun (semantik) bağlılığı kriterilərinin üzərində qurulur.

Belə ki, semantik modeldə hər hansı obyektə aid olan informasiya həmin obyekt haqqında biliklər toplusu kimi təsvir olunur. Semantik modelə uyğun olaraq, informasiya sistemləri, eləcə də web-xidmət təsvir olunan obyektlər haqqında biliklərə malik olur və bu bilikləri onlara mürciət edən istifadəçilərə (agentlərə) təqdim edir.

Semantik şəbəkələrdə biliklər strukturlaşdırılmış informasiya və nəticəçıxarma qaydaları çoxluqları şəklində təşkil olunur. Bu çoxluqlar avtomatik fikir yürütmələr üçün istifadə olunur.

Qeyd olunduğu kimi, hazırda internetin əsas xidməti olan web-texnologiya informasiyanın insanlara təqdim olunması məqsədini daşıyır və ona görə də onun ehtiyatları yalnız insanlar tərəfindən effektiv istifadə oluna bilir. Lakin semantik şəbəkənin əsasını təşkil edən semantik web internetin gələcək inkişafını istiqamətləndirərək, informasiya ehtiyatlarının in-

sanlarla yanaşı kompüterlər tərəfindən də emal olunmasına imkan verən formada təsvir olunması üçün müvafiq dillərin tətbiqini təmin etməlidir. Burada əsas məqsəd məlumatları (bilikləri) və onlar haqqında fikiryürütmə (nəticəçıxarma) qaydalarını təsvir etməyə imkan verən belə dilin yaradılmasından ibarətdir. Bu dil hər hansı biliklərin təqdim edilməsi üçün sistemdə mövcud olan nəticəçıxarma qaydalarını həmin sistemdə istifadə etməklə yanaşı, onları şəbəkə vasitəsilə digər analogi sistemlərə ötürməyə imkan verməlidir.

Semantik şəbəkənin əsas vəzifəsi şəbəkəyə çox güclü məntiqin gətirilməsi, yəni güclü məntiqə malik şəbəkənin yaradılmasıdır. Bu məntiq, obyektlərin mürəkkəb xassələrini təsvir etməyə imkan verməli, lakin eyni zamanda gözlənilməz sorğular formalaşdırmaqla axtarış agentini çıxılmaz vəziyyətə salmamalıdır.

Beləliklə, semantik web – yalnız insanlar üçün deyil, həm də informasiya ehtiyatlarının kompüterlər tərəfindən intellektual emalı üçün nəzərdə tutulmuş web-texnologiyadır. Semantik web kompüterlərə internetin informasiya ehtiyatlarını daha yaxşı anlamaq və onların insanlar tərəfindən istifadəsi üçün daha faydalı etmək imkanı verən texnologiyalar kompleksidir.

Semantik web-texnologiyanın reallaşdırılması üçün nəzərdə tutulan dil (nişanqoyma dili) kifayət qədər informativ olmalıdır ki, kompüter daxil edilmiş ifadələrin mənalarını müstəqil olaraq başa düşə bilsin. Bu cür dillər web-xidmətlərin axtarılıb tapılması, yüklənməsi (yerinə yetirilməsi), bir neçə web-xidmətin birgə istifadəsi, web-xidmət yükləndikdən sonrakı vəziyyətin nəzarət olunması (monitorinq) kimi tələblərə cavab verməlidirlər.

Semantik web şəbəkədə olan sənədlərin təsvir olunması üçün onları təşkil edən elementlərin (əlamətlər, göstəricilər, terminlər və s.) həmin sənədlərə münasibətləri, eləcə də bu

elementlər arasındakı əlaqələri modelləşdirilməlidir. Semantik şəbəkədə sənədlərin təsvir edilməsi, daha doğrusu onlar haqqında biliklərin təqdim olunması üçün aşağıdakı şəkildə üçlükdən ibarət olan çox sadə model reallaşdırılır:

$P(S,O) \Rightarrow (Subyekt, Predikat, Obyekt),$

burada P – predikat, S – subyekt və O – obyektidir.

Bu üçlüyün köməyi ilə sənədlər müəyyən subyektlə (şəxs, web-səhifə, sənəd, termin və s.) hər hansı obyekt (şəxs, web-səhifə, sənəd, termin və s.) arasında mövcud münasibətlərin (müəllif olmaq, istinad etmək, relevant olmaq və s.) təsviri şəkildə təqdim olunur. Burada subyekt və obyekt qismində istənilən ifadə və ya mürəkkəb strukturlar, predikat qismində isə bu obyektlər arasında mümkün ola biləcək münasibətlər istifadə edilir. Məsələn, $P(S,O)$ modelinə əsasən "Semantik şəbəkə web-texnologiyanın davamıdır" fikri aşağıdakı kimi təsvir oluna bilər:

Davamı olmaq (Semantik şəbəkə, web-texnologiya).

Müasir semantik şəbəkədə tətbiq edilən və şəbəkənin əsas informasiya modeli olan RDF (Resource Description Framework – ehtiyatların təsviri şablonu) dili də məlumatların təsvir olunması üçün mahiyyətə yuxarıda qeyd olunan modelə analoji olan müvafiq üçlükdən istifadə edilir. Burada subyekt və obyekt URI (Uniform Resource Identifier – ehtiyatın eyniləşdirmə identifikatoru) vasitəsi ilə verilir.

Qeyd olunmalıdır ki, hazırda semantik şəbəkə qarşısında duran məsələlərin həyata keçirilməsi istiqamətində bir sıra yeni texnologiyalar işlənilib hazırlanmış və istifadə olunmağa başlanmışdır: XML (eXtensible Markup Language – genişləndirilmiş nişanlama dili), OWL (Web Ontology Language – şəbəkə ontologiyaları dili), XHTML (eXtensible HTML – genişlənmiş HTML), XLST (Extensible Stylesheet Language Transformations – dillərin çevrilməsi üçün genişlənə bilən stil

cədvəlləri) və DAML-S (DARPA Agent Markup Language – DARPA web-xidmətlər üzrə agentlərin nişanlanması dili).

Semantik şəbəkələr hazırda inkişaf mərhələsindədir. Onların inkişafı informasiya sistemlərinin, eləcə də informasiya-axtarış sistemlərinin intellektuallığının artırılmasını təmin edəcəkdir.

6.8. Qeyri-səlis model

Effektiv İAS-ın yaradılması üçün axtarış aparılan informasiya mühitini bütövlükdə öyrənmək, informasiya ehtiyatlarının tematik bölünməsi, indeksləşdirilməsi və axtarılması, sənədlərin və sorğuların təqdim olunması üsullarını işləyib hazırlamaq tələb olunur. İnformasiya axtarışının qeyri-səlis çoxluqlar və münasibətlər nəzəriyyəsinə əsaslanan modelində sənədlər, tematik kataloqlar və terminlər çoxluqları, eləcə də bu çoxluqlar arasındakı münasibətlər öz əksini tapmalıdır.

Qeyri-səlis model istifadəçi sorğularına daha relevant olan sənədləri tapmağa və tematik profili müəyyənləşdirməyə, habelə terminlərin sinonimlərindən və assosiativ sözlərdən ibarət lüğətləri istifadə etməklə avtomatik olaraq tematik kataloqları yaratmağa və mövcud tematik kataloqları təkmilləşdirməyə imkan verir. Bu məqsədlə modeldə qeyri-səlis çoxluqlar nəzəriyyəsindən, o cümlədən Belman-Zadə yanaşmasından istifadə edilir.

Tutaq ki, $D = \{d_i\}_I$ – sənədlər çoxluğuudur. Bu çoxluğu kitabxana-informasiya sistemlərinə analogi olaraq mövzular üzrə müvafiq altçoxluqlara (tematik kataloqlara) bölmək tələb olunur.

Burada tematik kataloqlar dedikdə tematik profilləri ilə müəyyən olunan təsnifat istiqamətləri çoxluğu başa dü-

şülür. Hər bir tematik kataloq özünün profilinə uyğun olaraq, məxsusi deskriptorlar, yəni terminlər, açar sözləri və ya digər leksik vahidlər çoxluqları ilə xarakterizə olunur.

Tutaq ki, $K = \{K_l\}_L$ – tematik kataloqlar çoxluğu, $T = \{t_j\}_J$ – sistemdə istifadə olunan bütün terminlər çoxluğu, T_i^K – K_l profilinin tematik istiqamətini xarakterizə edən terminlər çoxluğudur. Qeyd olunmalıdır ki, sonuncu çoxluqlar öz aralarında qismən kəsişə bilər:

$$\forall i \neq j \text{ üçün } T_i^K \cap T_j^K \neq \emptyset. \quad (6.9)$$

Aydın ki, sistemdə istifadə olunan bütün tematik profillərin terminləri çoxluqları birlikdə axtarış sisteminin ümumi terminləri çoxluğunu təşkil edir:

$$T = \bigcup_{l=1}^L T_l^K. \quad (6.10)$$

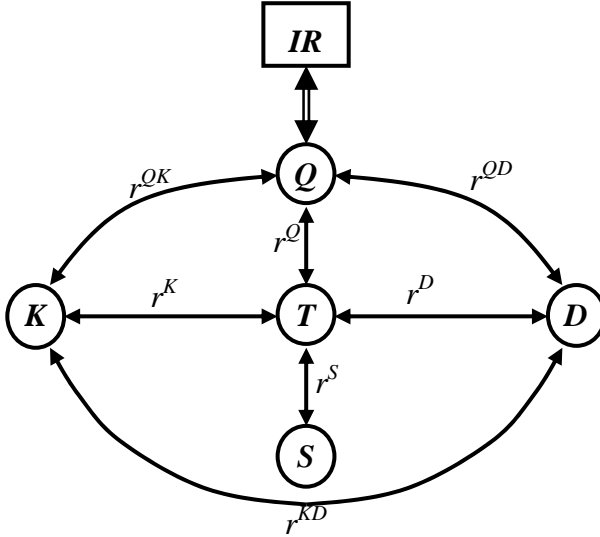
Sadəlik üçün fərz edək ki, sistemin terminlər çoxluğuna daxil olan $t_j \in T, j = \overline{1, J}$ terminlərinin sinonimləri və assosiativ sözləri vahid çoxluq təşkil edir. Elementləri arasındakı münasibətlər qeyri-səlis münasibətlər şəklində müəyyən olunan sinonimlər və assosiativ sözlər çoxluğunu $S = \{s_v\}_V$ kimi işarə edək.

Yuxarıda qeyd olunanları nəzərə alaraq, informasiya axtarışını aşağıdakı şəkildə təqdim etmək olar:

$$I_R = \{K, D, T, Q, R\}, \quad (6.11)$$

burada I_R – informasiya axtarışının istifadəçiyə təqdim olunan nəticəsi, Q – istifadəçi sorğusu, $R = \{R^D, R^K, R^S, R^Q\}$ – qeyri-səlis münasibətlər çoxluqları çoxluğudur. Bu çoxluq “sənəd – termin” (R^D), “tematik kataloq – termin” (R^K), “termin – sinonim” (R^S) və “sorğu – termin” (R^Q) şəklində müəyyən edilən münasibətləri özündə birləşdirir.

Şəkil 6.1-də informasiya axtarışının qeyri-səlis modeli istiqamətlənmiş qraf şəklində təsvir olunmuşdur. Bu qrafın təpələri yuxarıda qeyd olunan çoxluqları, qövsləri isə bu çoxluqlar arasındakı münasibətləri göstərir. Bu qrafda informasiya axtarışının qeyri-səlis modelinə daxil olan bütün çoxluqlar və münasibətlər öz əksini tapmışdır.



Şək.6.1. İnformasiya axtarışının qeyri-səlis modeli istiqamətlənmiş qraf şəklində təsviri

Tematik profillər və terminlər çoxluqları arasındakı münasibətləri müəyyən edən çoxluğu tematik rubrikator adlandıracağıq. Qeyd olunmalıdır ki, axtarış sistemlərində tematik rubrikator qismində istənilən universal bibliografik təsnifatlandırıcı, məsələn, UOT – universal onluq təsnifatlandırıcı, KBT – kitabxana bibliograf təsnifatlandırıcı və s. istifadə oluna bilər. Terminlər çoxluğunun yaradılması üçün digər yanaşma da mümkündür. Bu yanaşmanın məğzi terminlər çoxluğunun indekslər və ya meta-verilənlər baza-

ları şəklində axtarış sisteminin özü tərəfindən fəaliyyət prosesində yaradılmasından və doldurulmasından ibarətdir.

Yuxarıda qeyd olunduğu kimi, hər bir tematik profil müəyyən terminlər çoxluğu ilə təyin edilir. Məlumdur ki, bəzi terminlər yalnız bir tematik profilə, digər terminlər isə bir neçə tematik profilə aid (relevant) ola bilər. Burada deyirlər ki, bu terminlər verilmiş tematik profili təsvir edirlər.

Tematik rubrikator $L \times J$ ölçülü qeyri-səlis relyasion matris şəklində təqdim olunur. Bu matrisin sətirləri tematik profillərə, sütunları isə terminlərə uyğun olur.

Beləliklə, tematik rubrikatoru aşağıdakı kimi təqdim etmək olar:

$$K_l = \{t_j / \varphi_{t_j}(K_l)\}, \quad l = \overline{1, L}, \quad j = \overline{1, J}. \quad (6.12)$$

Burada $\varphi_{t_j}(K_l) - t_j$ termininin K_l tematik kataloquna relevantlığı (daxil olması) dərəcəsini müəyyən edən funksiyadır. Başqa sözlə, $\varphi_{t_j}(K_l)$ terminin tematik profilə daxilolma (mənsubiyyət) funksiyasıdır.

Tematik profillər çoxluqları ilə terminlər çoxluqları arasındakı R^K münasibəti aşağıdakı kimi müəyyən edilir:

$$R^K = \{r_l^K\}_L, \\ r_l^K = \{\varphi_{t_j}(K_l) : T \times T^K \rightarrow [0,1]\}, \quad l = \overline{1, L}, \quad j = \overline{1, J}. \quad (6.13)$$

R^S və R^K matrislərinin elementlərinin qiymətləri ilkin mərhələdə İAS-ın yaradılması zamanı müəyyən edilir. Bunun üçün ekspert qiymətləndirilməsi üsulundan istifadə oluna bilər. Sistemin fəaliyyəti prosesində $\varphi_{t_j}(K_l)$ funksiyasının qiymətləri adaptasiya olunur, yəni korrektə edilir.

İndi isə indeksləşdirilmiş sənədlərin qeyri-səlis modelin köməyi ilə axtarış sisteminin indekslər bazasında təqdim olunması prinsiplərinə baxaq. Yuxarıda qeyd olunduğu

kimi, indeksləşdirmə prosesinin nəticəsində informasiya fəzasının hər bir sənədinə qarşı T çoxluğundan bir sıra terminlər, yəni onun alt çoxluğu qoyulur. Terminlərə onların hər bir sənəd üçün vaciblik dərəcələrini göstərən çəki əmsalları mənimsədilir.

d_i sənədinin terminləri (açar sözləri) çoxluğunu T_i^D (qeyri-səlis çoxluqdur) ilə işarə edək. Hər bir sənədə münasibətdə terminlərin çəki əmsalları bu terminlərin T_i^D çoxluğuna daxilolma funksiyası ilə müəyyən olunur. Bu funksiya $[0,1]$ intervalında qiymətlər alır. Onun qiyməti terminin çoxluğa daxilolma dərəcəsini göstərir. Belə ki, 1 qiyməti terminin çoxluğa tamamilə daxil olduğunu, 0 qiyməti isə onun bu çoxluğa aid olmadığını müəyyən edir.

Qeyd olunmalıdır ki, bütün T_i^D altçoxluqları normallaşdırılır, yəni T çoxluğunun bu altçoxluqlarda olmayan terminləri onlara əlavə edilir, lakin həmin terminlərin daxilolma funksiyalarının qiymətləri sıfır qəbul olunur.

Beləliklə, D sənədlər çoxluğunu $I \times J$ ölçülü qeyri-səlis relyasion matris şəklində təqdim etmək olar. i sətiri ilə j sütununun kəsişməsində yerləşən element t_j termininin d_i sənədi üçün çəkisini müəyyən edir:

$$d_i = \{t_j / \omega_{i,j}(d_i)\}, \quad i = \overline{1, I}, \quad j = \overline{1, J}. \quad (6.14)$$

Eyni zamanda t_j termini ilə d_i sənədi arasındakı münasibət aşağıdakı kimi təyin edilir:

$$R^D = \{R_i^D\}_I, \\ R_i^D = \{\omega_{i,j}(d_i) : T \times T^D \rightarrow [0,1]\}, \quad i = \overline{1, I}, \quad j = \overline{1, J}, \quad (6.15)$$

burada $\omega_{i,j}(d_i) - t_j$ terminin T_i^D terminlər çoxluğuna daxilolma funksiyasıdır. Onun qiyməti sənədlərin indeksləşdirilməsi zamanı hesablanır. R^D münasibətlər çoxluğunu çox vaxt W ilə işarə edirlər.

VII FƏSİL

QEYRİ-SƏLİS MODEL ƏSASINDA İNFORMASIYA AXTARIŞI ÜSULLARI

**Sənədlərin profillərinə görə informasiya fəzasının
tematik kataloqlara bölünməsi**

**Tematik profilin seçilməsinin dəqiqliyinin artırılması
və tematik kataloqun keyfiyyətinin yaxşılaşdırılması
üsulları**

**İstifadəçinin qeyri-səlis sorğusuna əsasən daha
relevant sənədlərin axtarılması algoritmi**

**Qeyri-səlis üstünlük münasibətləri əsasında relevant
informasiyanın axtarılması**

**Qeyri-səlis üstünlük münasibətləri əsasında paylanmış
informasiya axtarışının təşkili**

7.1. Sənədlərin profillərinə görə informasiya fəzasının tematik kataloqlara bölünməsi

İnternetin informasiya fəzasının sənədlərinin avtomatik olaraq tematik istiqamətlər üzrə ayrılması, yəni onların tematik kataloqlara bölünməsi məsələsi kitabxana-informasiya sistemlərində tematik kataloqların avtomatik yaradılması məsələsinə analogiya təşkil edir.

Əgər $T_i^D - d_i$ sənədinin terminləri çoxluğu və $T - əvvəlki$ fəsildə müəyyən edilmiş tematik rubrikatordursa, onda d_i sənədinin rubrikatora daxil olan tematik kataloqlara relevantliq dərəcələrini bu çoxluqların kəsişməsi şəklində tapmaq olar:

$$T \cap T_i^D = \left(\bigcup_{l=1}^L T_l^K \right) \cap T_i^D = \bigcup_{l=1}^L (T_l^K \cap T_i^D). \quad (7.1)$$

d_i sənədi üçün daha relevant tematik kataloqun aşkar edilməsi üçün Belman-Zadə yanaşmasından istifadə etmək olar. Bu proses mərhələlərlə həyata keçirilir.

Birinci mərhələdə d_i sənədinin K çoxluğunun hər bir tematik kataloquna relevantliq dərəcəsi müəyyən olunur. Yuxarıda qeyd olunduğu kimi, d_i sənədi sistemdə T çoxluğuna daxil olan terminlərlə həmin sənəd arasındakı R_i^D münasibəti ilə təqdim olunur:

$$R_i^D = \{ \omega_{i1}, \omega_{i2}, \dots, \omega_{iI} \}, i = \overline{1, I}, \quad (7.2)$$

burada $\omega_{ij} = \omega_{i_j}(d_i)$.

Tematik kataloqlar da analoji qaydada təqdim olunurlar:

$$R_l^K = \{ \varphi_{l1}, \varphi_{l2}, \dots, \varphi_{lL} \}, l = \overline{1, L}, \quad (7.3)$$

burada $\varphi_{ij} = \varphi_{i_j}(K_l)$.

Onda d_i sənədinin K_l tematik kataloquna relevantlıq münasibətini R_i^D və R_l^K münasibətləri çoxluqlarının kəsişməsi kimi vermək olar, yəni:

$$R_{il}^{DK} = R_i^D \cap R_l^K. \quad (7.4)$$

Məlum olduğu kimi, qeyri-səlis çoxluqların kəsişməsi bu çoxluqların uyğun elementlərinin cəbri hasili kimi müəyyən olunur:

$$\eta_{ij}^l = \omega_{ij} \cdot \phi_{lj}, \quad (7.5)$$

burada $\eta_{ij}^l \in R_{il}^{DK}$, $l = \overline{1, L}$, $j = \overline{1, J}$ – t_j termininə münasibətdə d_i sənədinin tematikasının K_l tematik kataloqunun profilinə relevantlıq dərəcəsidir.

İkinci mərhələdə sənədin tematikası üçün daha üstün olan (üstünlük təşkil edən) tematik kataloq tapılır. Tutaq ki, K^{ab} – d_i sənədi üçün müəyyən olunmuş abstrakt tematik kataloqdur. O, d_i sənədinin bütün terminləri üzrə mövcud tematik kataloqların müvafiq profilə ən yaxşı relevantlıq münasibətlərini özündə birləşdirən tematik kataloqdur.

K^{ab} tematik kataloqu $l = \overline{1, L}$ üçün bütün R_{il}^{DK} qeyri-səlis çoxluqlarının birləşməsi, yəni bütün terminlər üzrə hər bir kataloqlar üçün η_{ij}^l kəmiyyətləri arasından maksimumların tapılması yolu ilə müəyyən olunur:

$$\eta_{ij}^{ab} = \max_{l=1, L} \{ \eta_{ij}^l \}, \quad j = \overline{1, J}, \quad (7.6)$$

burada η_{ij}^{ab} – t_j termininə görə d_i sənədinin K^{ab} tematik kataloquna relevantlıq dərəcəsidir. Buna əsaslanaraq de-

mək olar ki, K^{ab} tematik kataloqu d_i sənədinə ən relevant olan abstrakt tematik kataloq olacaqdır.

İndi isə $\{K_l\}_L$ çoxluğundan d_i sənədinin bütün t_j terminlərini nəzərə almaqla tematik profili K^{ab} abstrakt kataloqunun profilinə ən yaxın olan K_l^* tematik kataloqunu tapaq. Bu məqsədlə $\{K_l\}_L$ çoxluğunun bütün K_l tematik kataloqlarının relevantlıq əmsallarının K^{ab} abstrakt kataloqunun relevantlıq əmsallarından ümumi orta kvadratik uzaqlaşmalarının nəzərə alınması tələb olunur. Orta kvadratik uzaqlaşmanı aşağıdakı düsturun köməyi ilə hesablamaq olar:

$$\lambda_{il} = \frac{\bar{J}}{J} \sum_{j=1}^J (\eta_{ij}^{ab} - \eta_{ij}^l)^2, \quad l = \overline{1, L}, \quad (7.7)$$

burada \bar{J} – d_i sənədinin K_l tematik kataloquna münasibətdə çəki əmsallarının qiyməti sıfır olmayan terminlərinin sayıdır.

Aydındır ki, çəki əmsalı minimal orta kvadratik uzaqlaşmaya malik olan tematik kataloq verilmiş sənəd üçün ən üstün tematik kataloq olacaqdır. Başqa sözlə:

$$\lambda_i^* = \min_{l=1, L} \{\lambda_{il}\}. \quad (7.8)$$

Beləliklə, λ_i^* minimal orta kvadratik uzaqlaşmaya malik olan K_l^* tematik kataloqu $\{K_l\}_L$ çoxluğunda ən yaxşı (ən relevant) tematik kataloqdur, yəni K_l^* tematik kataloqunun profili K^{ab} abstrakt kataloqunun profilinə ən çox relevantlıq təşkil edir. Deməli, K_l^* tematik kataloqunun profili verilmiş

d_i sənədinin tematikasına ən yaxın olan kataloqdur. Buradan belə nəticəyə gəlmək olar ki, d_i sənədini K_l^* tematik kataloquna daxil etmək lazımdır.

Əvvəlki fəsildə verilmiş mülahizələrə əsasən demək olar ki, hər bir sənəd yalnız bir tematik kataloqa deyil, bir neçə tematik kataloqa relevant ola bilər. Ona görə də tematikası verilmiş sənədin tematikasına müəyyən edilmiş dərəcədən az olmayaraq relevant olan bütün tematik kataloqların tapılması məsələsi aktualıq kəsb edir. Bu məqsədlə tematik kataloqların relevantlıq dərəcələri üçün $\varepsilon \in [0,1]$ sərhəd qiymətini daxil edək.

Beləliklə, yuxarıda qeyd olunan məsələni aşağıdakı şəkildə ifadə etmək olar: verilmiş sənədə relevantlıq dərəcəsinin λ_{il} orta kvadratik uzaqlaşmaları ε sərhəd qiymətdən kiçik olan daha üstün tematik kataloqları tapmalı.

Deməli, daha üstün tematik kataloqların müəyyənləşdirilməsi üçün (7.6) düsturundan λ_{il} orta kvadratik uzaqlaşmaları verilmiş sərhəd qiymətlərini aşmayan, yəni aşağıdakı şərti ödəyən bütün tematik kataloqları tapmaq tələb olunur:

$$\lambda_{il} \leq \varepsilon_l, i = \overline{1, I}, l = \overline{1, L}. \quad (7.9)$$

Təklif olunan üsulu əyani nümayiş etdirmək və onun adekvatlığını sübut etmək üçün aşağıda konkret nümunəyə baxaq.

Tutaq ki, terminlər çoxluğu on termindən ($J=10$), tematik rubrikator isə dörd tematik kataloqdan ($L=4$) ibarətdir, yəni:

$$T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}\}$$

və

$$K = \{K_1, K_2, K_3, K_4\}.$$

$K_l, l = \overline{1, 4}$ tematik kataloqlarının $t_j, j = \overline{1, 10}$ terminlərinə münasibəti R_l^K qeyri-səlis relyasion cədvəli ilə verilmişdir:

$$\begin{aligned}
K_1(t_j / \varphi_{1j}) &= \{t_1 / 0.75; t_2 / 0; t_3 / 0.63; t_4 / 0; \\
&\quad t_5 / 0.95; t_6 / 0.82; t_7 / 0; t_8 / 0; t_9 / 0.78; t_{10} / 0\}, \\
K_2(t_j / \varphi_{2j}) &= \{t_1 / 0; t_2 / 0.9; t_3 / 0.79; t_4 / 0.65; \\
&\quad t_5 / 0.6; t_6 / 0; t_7 / 0.9; t_8 / 0; t_9 / 0.69; t_{10} / 0.78\}, \\
K_3(t_j / \varphi_{3j}) &= \{t_1 / 0.56; t_2 / 0.83; t_3 / 0; t_4 / 0; \\
&\quad t_5 / 0.64; t_6 / 0; t_7 / 0.76; t_8 / 0.95; t_9 / 0; t_{10} / 0.69\}, \\
K_4(t_j / \varphi_{4j}) &= \{t_1 / 0.9; t_2 / 0.78; t_3 / 0.68; t_4 / 0.54; \\
&\quad t_5 / 0.8; t_6 / 0.2; t_7 / 0.86; t_8 / 0; t_9 / 0.71; t_{10} / 0.9\}.
\end{aligned}$$

Fərz edək ki, axtarış sistemi tərəfindən d_I sənədinin indeksləşdirilməsi nəticəsində terminlərin bu sənəd üçün aşağıdakı vaciblik dərəcələri müəyyən edilmişdir:

$$\begin{aligned}
d_1(t_j / \omega_{1j}) &= \{t_1 / 0; t_2 / 0.9; t_3 / 0.71; t_4 / 0.87; \\
&\quad t_5 / 0.54; t_6 / 0; t_7 / 0.6; t_8 / 0; t_9 / 0.95; t_{10} / 0.58\}.
\end{aligned}$$

Beləliklə, tematik kataloqlarla terminlər arasındakı münasibətlər

$$R_i^K = \begin{pmatrix} 0.75 & 0.9 & 0.63 & 0 & 0.95 & 0.82 & 0 & 0 & 0.78 & 0 \\ 0 & 0.8 & 0.79 & 0.65 & 0.6 & 0 & 0.9 & 0 & 0.69 & 0.78 \\ 0.56 & 0.83 & 0 & 0 & 0.64 & 0 & 0.76 & 0.95 & 0 & 0.69 \\ 0 & 0.78 & 0.68 & 0.54 & 0.8 & 0.2 & 0.86 & 0 & 0.71 & 0.9 \end{pmatrix}$$

matrisi vasitəsilə, sənədlərin terminlərlə təsvirini isə

$$R_1^D = (0 \quad 0.9 \quad 0.71 \quad 0.87 \quad 0.54 \quad 0 \quad 0.6 \quad 0 \quad 0.95 \quad 0.58).$$

vektoru şəklində təqdim etmək olar.

İndi isə (7.5) düsturuna əsasən η_{1j}^I əmsallarının qiymətlərini tapaq:

$$\eta_{ij}^l = \omega_{ij} \cdot \varphi_{lj}, \quad l = \overline{1,4}, \quad j = \overline{1,10}.$$

Nəticədə aşağıdakı cədvəli almış olarıq:

$$R_{jl}^{DK} = \begin{pmatrix} 0 & 0 & 0.45 & 0 & 0.51 & 0 & 0 & 0 & 0.74 & 0 \\ 0 & 0.81 & 0.56 & 0.57 & 0.32 & 0 & 0.54 & 0 & 0.66 & 0.45 \\ 0 & 0.75 & 0 & 0 & 0.35 & 0 & 0.46 & 0 & 0 & 0.4 \\ 0 & 0.68 & 0.48 & 0.47 & 0.43 & 0 & 0.52 & 0 & 0.67 & 0.52 \end{pmatrix}$$

(7.6) düsturuna əsasən K^{ab} abstrakt kataloqunun çəki əmsallarının qiymətlərini tapaq:

$$\eta_{1j}^{ab} = \max_{l=1,4} \{ \eta_{1j}^l \}, \quad j = \overline{1,10}.$$

$$\{ \eta_{1j}^{ab} \} = \{ 0 \quad 0.81 \quad 0.56 \quad 0.57 \quad 0.51 \quad 0 \quad 0.54 \quad 0 \quad 0.74 \quad 0.52 \}$$

Nəhayət, bütün tematik kataloqlar üçün (7.7) düsturuna əsasən orta kvadratik uzaqlaşma əmsallarını hesablayaq:

$$\lambda_{11} = \frac{5}{10} \cdot \sum_{j=1}^{10} (\eta_{1j}^{ab} - \eta_{1j}^1)^2 = 0.78,$$

$$\lambda_{12} = \frac{6}{10} \cdot \sum_{j=1}^{10} (\eta_{1j}^{ab} - \eta_{1j}^2)^2 = 0.03,$$

$$\lambda_{13} = \frac{6}{10} \cdot \sum_{j=1}^{10} (\eta_{1j}^{ab} - \eta_{1j}^3)^2 = 0.74,$$

$$\lambda_{14} = \frac{7}{10} \cdot \sum_{j=1}^{10} (\eta_{1j}^{ab} - \eta_{1j}^4)^2 = 0.04.$$

Tematikası K^{ab} abstrakt kataloqunun tematikasından ən az uzaqlaşan tematik kataloq K_2 -dir:

$$\lambda_1^* = \min_{l=1,4} \{\lambda_{1l}\} = \min \{\lambda_{11}, \lambda_{12}, \lambda_{13}, \lambda_{14}\} = \lambda_{12} = 0.03.$$

Bu o deməkdir ki, K_2 tematik kataloqu d_1 sənədinin tematikasına ən yaxın olan tematik kataloqdur və ona görə də d_1 sənədini K_2 tematik kataloquna daxil etmək lazımdır.

Əgər $K_i, i = \overline{1,4}$ tematik kataloqlarının K^{ab} abstrakt tematik kataloqundan uzaqlaşması üçün sərhəd qiymətini $\varepsilon = 0.1$ qəbul etsək, onda d sənədinə daha relevant tematik kataloqlar yalnız K_2 və K_4 olar, yəni:

$$\lambda_{12} < \lambda_{14} \leq 0.1.$$

Nəticədə d_1 sənədi K_2 və K_4 tematik kataloqlarının hər ikisinə daxil edilməlidir.

7.2. Tematik profilin seçilməsinin dəqiqliyinin artırılması və tematik kataloqun keyfiyyətinin yaxşılaşdırılması üsulları

Yuxarıda qeyd olunduğu kimi, axtarışın dolğunluğunu və dəqiqliyini yaxşılaşdırmaq üçün sinonimlər, assosiativ sözlər çoxluqlarından və digər lüğətlərdən istifadə etmək olar. Həmçinin qəbul etmişdik ki, terminlərin sinonimləri və assosiativ sözləri bir çoxluğun tərkibində birləşirlər, yəni

$$S = \{s_v\}_V = \{S_j\}_J, \quad (7.10)$$

burada $S_j - t_j$ terminlərinin sinonimləri və assosiativ sözləri çoxluğuudur. Rahatlıq üçün gələcəkdə S çoxluğunu sinonimlər çoxluğu adlandıracağıq.

Qeyd olunmalıdır ki, sinonimlər, eyni zamanda, həm də termin rolunda çıxış edə bilər. Təbii ki, istənilən açar söz müəyyən sənədin hər hansı termini üçün sinonim olmaqla yanaşı, eyni zamanda, digər sənəd üçün vacib (əhəmiyyətli) termin ola bilər. Bu o deməkdir ki, istənilən $t_j \in T$ termini üçün onun S_j sinonimləri çoxluğu sistemin T terminləri çoxluğunun alt çoxluğudur. Buradan çıxır ki, S çoxluğu T çoxluğunun alt çoxluğudur. Ona görə də rahatlıq naminə gələcəkdə S çoxluğu əvəzinə T çoxluğunu istifadə edəcəyik. Beləliklə, istənilən $j, v = \overline{1, J}$ üçün t_j termini ilə s_v sinonimi arasındakı münasibətlər əvəzinə t_j və t_v terminləri arasındakı münasibətlər baxılacaqdır.

t_j və t_v terminləri arasındakı münasibəti, yəni t_v termininin t_j termininə yaxınlıq dərəcəsini göstərən əmsali r_{jv}^S ilə işarə edək. r_{jv}^S münasibətləri $J \times J$ ölçülü qeyri-səlis relyasion matris şəklində verilir. Onun elementləri aşağıdakı qeyri-səlis münasibət kimi təyin edilir:

$$r_{jv}^S = \{v_{jv} : T \times T \rightarrow [0,1]\}, j, v = \overline{1, J}, \quad (7.11)$$

burada $v_{jv} = t_v$ termininin S_j sinonimlər (t_j termininin sinonimləri) çoxluğuna daxilolma funksiyası, başqa sözlə, t_j və t_v terminlərinin yaxınlıq (sinonimlik) dərəcəsidir. Əgər $t_j \in T$ və $t_v \in T$ terminləri üçün $t_v \in S_j$ şərti ödənilməzsə, onda $v_{jv} = 0$ qəbul olunur.

Sinonimlər və onların terminlərə mənaca yaxınlıq dərəcələri haqqında məlumatlar sənədin tematikası haqqında bilikləri genişləndirməyə kömək edir. Bu isə öz növbəsində onun tematik profilini daha dəqiq müəyyənləşdirməyə və ona daha relevant olan tematik kataloqun seçilməsinə

imkan verir. Sənədlər haqqında biliklərin təkmilləşdirilməsi (tamamlanması) üçün onun terminləri və bu terminlərin sinonimləri haqqında bilikləri birləşdirmək lazımdır. Əgər T_i^D - d_i sənədin terminləri çoxluğu və S_j isə t_j termininin sinonimləri çoxluğu varsa, onda

$$\tilde{T}_i^D = \bigcup_{t_j \in T_i^D} (S_j \cap T_i^D), \quad (7.12)$$

çoxluğu verilmiş sənəd üçün təkmilləşdirilmiş terminlər çoxluğu olacaqdır.

Yeni \tilde{T}_i^D çoxluğuna uyğun olan yeni $\tilde{\omega}_{ij}$ münasibətlərinin qiymətlərini aşağıdakı kimi hesablamaq olar:

$$\tilde{\omega}_{ij} = \max_{v=1, J} \{v_{jv} \cdot \omega_{iv}\}, j = \overline{1, J} \quad (7.13)$$

(7.13) düsturu bütün terminlərin çəki əmsallarının yeni (yaxşılaşdırılmış) qiymətlərini verir. Çəki əmsallarının ω_{ij} qiymətlərinin əvəzinə yeni $\tilde{\omega}_{ij}$ qiymətlərini istifadə etməklə (7.5)-(7.8) məsələsinin həlli daha effektiv nəticələrin əldə olunmasına imkan verir.

Təklif olunan üsulun adekvatlığını nümayiş etdirmək məqsədilə aşağıdakı nümunəyə baxaq.

Əvvəlki nümunədə müəyyən olunmuş həmin T və K çoxluqlarını, eləcə də R_l^K və R_l^D münasibətlərini götürək. Tutaq ki, bundan əlavə, $R^S = \{r_{jv}^{T}\}_{J \times J}$ münasibətlər matrisi verilmişdir:

$$R^S = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \\ S_{10} \end{pmatrix} = \begin{pmatrix} 1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0 & 1 & 0.9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 1 \end{pmatrix}$$

Onda (7.14) düsturuna əsasən aşağıdakı qiymətləri alırıq:

$$R_1^D = \{0.81 \ 0.9 \ 0.78 \ 0.87 \ 0.54 \ 0.86 \ 0.6 \ 0 \ 0.95 \ 0.86\}$$

Çəki əmsallarının $\tilde{\omega}_{ij}$ yeni qiymətlərini istifadə etməklə (7.5)-(7.8) düsturları vasitəsilə hesablamalar aparmaqla kataloqların tematikalarının abstrakt kataloqun tematikasından orta kvadratik uzaqlaşma əmsalının aşağıdakı qiymətlərini alırıq:

$$\lambda_{11} = 0.94,$$

$$\lambda_{12} = 0.64,$$

$$\lambda_{13} = 1.1,$$

$$\lambda_{14} = 0.56.$$

Buradan görünür ki, əvvəlki nümunədən fərqli olaraq, K_4 tematik kataloqu verilmiş sənədə daha relevantdir. Belə ki:

$$\lambda_{14} = \min_{l=1,4} \lambda_{1l} = 0.56.$$

Qeyd olunmalıdır ki, K_2 tematik kataloqunun orta kvadratik uzaqlaşma əmsalının qiyməti λ_{14} əmsalından çox cüzi böyükdür, ona görə də bu kataloqu da d_1 sənədinə relevant kataloq kimi qəbul etmək olar.

7.3. İstifadəçinin qeyri-səlis sorğusuna əsasən daha relevant sənədlərin axtarılması alqoritmı

Yuxarıda qeyd olunduğu kimi, istifadəçinin sorğuları axtarış əlamətlərini müəyyən edən açar sözlərdən və ya terminlərdən tərtib olunur. Əvvəlki paraqraflarda istifadəçi sorğusu $Q = \{t_h\}_H$ kimi işarə edilmişdi.

İndekslər bazasında sənədlərin təqdim olunması formatına uyğun olaraq, İAS axtarış prosesinin əvvəlində istifadəçi sorğularını da müvafiq formata çevirir. Sorğuya daxil edilmiş hər bir açar sözün və ya terminin həmin sorğuya münasibəti müəyyən edilir. Başqa sözlə, açar sözlərin tərtib edilmiş sorğu üçün vaciblik dərəcəsi göstərilir. Ümumiyyətlə, sorğu qeyri-səlis münasibət şəklində təqdim olunur:

$$Q = \{t_h / \alpha_{t_h}(Q)\}_h, \quad h = \overline{1, H} \quad (7.14)$$

"Sorğu-termin" münasibəti isə aşağıdakı kimi müəyyən edilir:

$$R^Q = \{r_h^Q\}_H = \{\alpha_{t_h}(Q): T^D \times T^Q \rightarrow [0,1]\}_h, \quad (7.15)$$

$$t_h \in T^Q, \quad h = \overline{1, H}$$

burada T^Q – Q sorğusunu təşkil edən terminlər çoxluğu, $\alpha_h = \alpha_{t_h}(Q) - t_h$ termininin Q sorğusu üçün vaciblik əmsalı, $r_h^Q - t_h$ termini ilə Q sorğusu arasındakı qeyri-səlis münasibətdir.

Bundan əlavə, İAS axtarış prosesinə başlamazdan qabaq T^Q çoxluğunu normallaşdırır, yəni T çoxluğunun T^Q çoxluğunda olmayan terminlər bu çoxluğa əlavə olunur, lakin onların vaciblik əmsallarının qiyməti sifıra bərabər qəbul edilir. Onda (7.15) düsturu aşağıdakı şəkli almış olur:

$$\begin{aligned} R^{QT} &= \left\{ \alpha_h^T : T \times T^Q \rightarrow [0,1] \right\}, \quad h = \overline{1, J}, \\ R^{QT} &= \left\{ r_h^{QT} \right\}_J, \end{aligned} \quad (7.16)$$

burada

$$\alpha_h^T = \begin{cases} \alpha_h, & t_h \in T^Q \text{ olarsa,} \\ 0, & t_h \notin T^Q \text{ olarsa.} \end{cases}$$

Axtarışın dolğunluğunu və dəqiqliyini yaxşılaşdırmaq üçün sorğunun terminlərinin sinonimləri çoxluğundan istifadə etmək olar:

$$\tilde{T}^Q = \bigcup_{t_h \in T^Q} (S_h \cap T^Q), \quad h = \overline{1, J}. \quad (7.17)$$

Uyğun olaraq,

$$\tilde{\alpha}_h = \max_{j=1, J} \{ v_{jh} \cdot \alpha_h^T \}, \quad h = \overline{1, J}. \quad (7.18)$$

Sonralar Q sorğusundan danışdıqda fərz edəcəyik ki, o, T^Q normallaşdırılmış və tamamlanmış qeyri-səlis çoxluqdur.

İndi isə belə qeyri-səlis Q sorğusuna əsasən informasiyanın axtarılması məsələsinə baxaq. Aydındır ki, axtarışın nəticəsi istifadəçi sorğusuna daha relevant olan sənədlərin siyahısından ibarət olacaqdır, yəni:

$$I_R = \{ d_g \}_G. \quad (7.19)$$

Belə ki,

$$I_R = T^Q \cap T^D = \bigcup_{i=1, I} (T^Q \cap T_i^D). \quad (7.20)$$

Q sorğusuna daha relevant sənədlərin müəyyənləşdirilməsi üçün (7.18) məsələsi (7.1) məsələsinə analogiya təşkil edir və onun həlli üçün də Belman-Zadə yanaşmasından istifadə etmək olar.

R_i^D və R^Q münasibətlər çoxluqlarının kəsişməsi d_i sənədinin Q sorğusuna relevantlıq dərəcəsini verir, yəni:

$$R_i^{DQ} = R_i^D \cap R^Q. \quad (7.21)$$

Əgər

$$R_i^{DQ} = \{\gamma_{ij}\}, \quad i = \overline{1, I}, \quad j = \overline{1, J} \quad (7.22)$$

olarsa, onda çoxluqların kəsişməsinin tapılması qaydasına uyğun olaraq,

$$\gamma_{ij} = \mu_{ij} \cdot \alpha_j, \quad i = \overline{1, I}, \quad j = \overline{1, J}, \quad (7.23)$$

burada $\gamma_{ij} - t_j$ termininə görə d_i sənədinin Q sorğusuna relevantlıq dərəcəsidir.

Tutaq ki, $d^{ab} - Q$ sorğusuna ideal dərəcədə relevant olan abstrakt sənəddir. Onda t_j termininə görə d^{ab} sənədinin Q sorğusuna relevantlıq dərəcəsini aşağıdakı kimi müəyyən etmək olar:

$$\gamma_j^{ab} = \max_{i=\overline{1, I}} \{\gamma_{ij}\}, \quad j = \overline{1, J}, \quad (7.24)$$

İndi isə hər bir sənəd üçün sorğunun bütün açar sözlərini nəzərə almaqla $\gamma_{ij}, i = \overline{1, I}, j = \overline{1, J}$ əmsallarının $\gamma_j^{ab}, j = \overline{1, J}$ əmsallarından ümumi orta kvadratik uzaqlaşmasını hesablayaq. Bunun üçün aşağıdakı düsturdan istifadə etmək olar:

$$\lambda_i = \frac{\bar{J}}{J} \sum_{j=1}^J (\gamma_j^{ab} - \gamma_{ij})^2, i = \overline{1, I}. \quad (7.25)$$

Sorğuya "ideal uyğun sənəddən" (abstrakt sənəddən) minimal uzaqlaşmaya malik olan sənəd ən relevant sənəd olacaqdır:

$$\lambda^* = \min_{i=1, I} \{\lambda_i\}. \quad (7.26)$$

(7.23)-(7.26) məsələsinin həlli yeganə ən relevant d^* sənədini tapmağa imkan verir. Lakin praktikada, adətən, yalnız bir ən relevant sənədi deyil, müəyyən relevantliq dərəcəsi olan bütün sənədləri tapmaq tələb olunur.

Bu məsələnin həlli üçün relevantliqın sərhəd qiymətini – ε parametrini daxil edək. Onda

$$\lambda_i \leq \varepsilon, i = \overline{1, I} \quad (7.28)$$

şərtlərini ödəyən bütün sənədləri verilmiş sorğuya relevant qəbul edəcəyik.

Aşağıdakı nümunənin köməyi ilə üsulun adekvatlığını göstərək.

Tutaq ki, $Q = \{t_3 / 0.95, t_5 / 0.8, t_9 / 0.8\}$ – istifadəçi sorğusu, $D = \{d_1, d_2, d_3, d_4, d_5\}$ isə informasiya fəzasının axtarış sistemi tərəfindən indeksləşdirilmiş sənədləri çoxluğudur. İndeksləşdirmə nəticəsində

$T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9, t_{10}\}$ çoxluğunun terminləri ilə d_j sənədləri arasında aşağıdakı münasibətlər aşkar olunmuşdur:

$$\begin{aligned}
d_1(t_j / \mu_{1j}) &= \{t_1 / 0.7; t_2 / 0; t_3 / 0.9; t_4 / 0.8; t_5 / 0; \\
&\quad t_6 / 0.7; t_7 / 0; t_8 / 0.8; t_9 / 0; t_{10} / 0.9\}^{\cdot} \\
d_2(t_j / \mu_{2j}) &= \{t_1 / 0.6; t_2 / 0; t_3 / 0.8; t_4 / 0; t_5 / 0.9; \\
&\quad t_6 / 0.9; t_7 / 0; t_8 / 0; t_9 / 0.8; t_{10} / 0\}^{\cdot} \\
d_3(t_j / \mu_{3j}) &= \{t_1 / 0; t_2 / 0.6; t_3 / 0.9; t_4 / 0; t_5 / 0.9; \\
&\quad t_6 / 0; t_7 / 0; t_8 / 0; t_9 / 0.7; t_{10} / 0\}^{\cdot} \\
d_4(t_j / \mu_{4j}) &= \{t_1 / 0; t_2 / 0; t_3 / 0; t_4 / 0; t_5 / 0.9; \\
&\quad t_6 / 0.7; t_7 / 0.8; t_8 / 0; t_9 / 0.6; t_{10} / 0\}^{\cdot} \\
d_5(t_j / \mu_{5j}) &= \{t_1 / 0.6; t_2 / 0.7; t_3 / 0.9; t_4 / 0; t_5 / 0; \\
&\quad t_6 / 0; t_7 / 0.6; t_8 / 0; t_9 / 0; t_{10} / 0.8\}^{\cdot}
\end{aligned}$$

T^Q çoxluğunu T çoxluğunun elementləri ilə aşağıdakı kimi tamamlamaq olar:

$$T^Q = \{t_1 / 0, t_2 / 0, t_3 / 0.95, t_4 / 0, t_5 / 0.8, t_6 / 0, t_7 / 0, t_8 / 0, t_9 / 0.8, t_{10} / 0\}.$$

(7.23) düsturuna əsasən γ_{ij} əmsallarını tapaq:

$$\left\{ \gamma_{ij} \right\} = \begin{pmatrix} 0 & 0 & 0.86 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.76 & 0 & 0.72 & 0 & 0 & 0 & 0.64 & 0 \\ 0 & 0 & 0.86 & 0 & 0.72 & 0 & 0 & 0 & 0.56 & 0 \\ 0 & 0 & 0 & 0 & 0.72 & 0 & 0 & 0 & 0.48 & 0 \\ 0 & 0 & 0.86 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Onda (7.24) düsturunun köməyi ilə γ_j^{ab} , $j = \overline{1, J}$ əmsalının qiymətini hesablayaq:

$$\{V_j^{ab}\} = (0 \ 0 \ 0.86 \ 0 \ 0.72 \ 0 \ 0 \ 0 \ 0.64 \ 0).$$

(7.25) düsturundan:

$$\{\lambda_i\} = \begin{bmatrix} 0.82 \\ 0.05 \\ 0.06 \\ 0.41 \\ 0.68 \end{bmatrix}$$

qiymətləri, (7.28) düsturundan isə

$$\lambda^* = 0.05$$

alınır. Buradan görünür ki, d_2 sənədi verilmiş sorğuya daha relevantdir. Əgər sərhəd qiymətlərini $\varepsilon = 0.1$ qəbul etsək, onda verilmiş D sənədlər çoxluğundan d_2 və d_3 sənədləri sorğuya relevant olar.

7.4. Qeyri-səlis üstünlük münasibətləri əsasında relevant informasiyanın axtarılması

Praktikada, adətən, informasiya axtarışı üçün istifadəçi tərəfindən formalaşdırılmış sorğular bir neçə açar sözündən (termindən) ibarət olur. Axtarış nəticəsində sistem istifadəçi sorğusuna əsasən müəyyən relevantlıq dərəcəsinə malik sənədlər çoxluğunu istifadəçiyə təqdim edir. Lakin qeyd olunmalıdır ki, tapılıb istifadəçiyə verilmiş sənədlər sorğunun ayrı-ayrı açar sözlərinə relevantlıq dərəcələrinə görə fərqlənilir.

Bundan əlavə, axtarış üçün sorğuya daxil edilmiş açar sözlərin vaciblik dərəcələri fərqli olur. Belə ki, istifadəçi-

yə məlumdur ki, onun sorğuya daxil etdiyi açar sözlərin hansı daha vacib, hansı isə nisbətən az vacibdir. Ona görə də onların vaciblik əmsalları sorğunun formalaşdırılması zamanı istifadəçi tərəfindən müəyyən edilir.

Yuxarıda qeyd olunanları nəzərə alaraq, axtarış sistemi tərəfindən tapılıb istifadəçiyə verilmiş sənədlər arasından sorğuya daha yüksək relevantlıq dərəcəsinə malik olan sənədlərin orta ölçülü seçimini aparmaq lazım gəlir. Aşağıda istifadəçi sorğusuna daha relevant olan sənədlərin tapılması üçün sənədlərin qeyri-səlis üstünlük münasibətlərinə əsaslanmış üsul baxılır.

Burada informasiya fəzasında relevant sənədlərin axtarılması məsələsi informasiya axtarışının qeyri-səlis modeli çərçivəsində daha münasib (relevant) informasiya ehtiyatının seçilməsi üzrə qərar qəbul etmə məsələsi kimi baxılır. Yuxarıda qeyd olunduğu kimi, bu model çərçivəsində informasiya fəzası sənədlər və axtarış terminləri, eləcə də onlar arasındakı qeyri-səlis münasibətlər çoxluqları vasitəsilə təqdim edilir.

Bu modelin gələcəkdə istifadə olunacaq bəzi məqamlarını burada dəqiqləşdirək. Əvvəlki fəsildə verilən təriflərə uyğun olaraq, $\omega_{ij} - t_j$ termininin d_i sənədinə daxilolma funksiyasıdır. O, terminlər və sənədlər arasındakı qeyri-səlis münasibətləri təsvir edir. Bu münasibətlər indeksləşdirmə üsullarının (avtomatik indeksləşdirmə zamanı) və ya müvafiq sahədə ekspertlərin biliklərinin (tematik kataloqlar olduqda) köməyi ilə müəyyən edilir.

Yuxarıda qeyd olunduğu kimi, sorğuya daxil edilmiş terminlər axtarılan sənədlər (onların tematikası) üçün müxtəlif vaciblik dərəcələrinə malik olurlar. Ona görə də sorğuların tərtibatı zamanı istifadəçi açar sözlərdən əlavə

onların bu sorğu üçün vaciblik dərəcələrini də müəyyən-
ləşdirir, yəni (7.14) və (7.15) düsturlarına uyğun olaraq
açar sözlərin tərtib olunan sorğu üçün vaciblik əmsallarını
da bu sorğuya daxil edir.

İndi isə qeyri-səlis üstünlük münasibətləri əsasında rele-
vant sənədlərin axtarılması məsələsinə baxaq. Tutaq ki, istifa-
dəçi sorğusu T terminlər çoxluğunun T^Q alt çoxluğundan ibar-
ətdir və bütün $t_h \in T^Q$ terminləri üçün $\alpha_{t_h}(Q) \rightarrow [0,1]$ funk-
siyası müəyyən olunmuşdur. Qeyd olunmalıdır ki, $T(t_h \in T)$
çoxluğunda olan, lakin $T^Q(t_h \notin T^Q)$ çoxluğuna daxil olma-
yan terminlərin vaciblik dərəcələri bu sənəd üçün sıfıra bəra-
bərdir. Başqa sözlə, T^Q çoxluğu T çoxluğunun digər termin-
ləri ilə tamamlanır və onların vaciblik əmsallarına sıfır qiymə-
ti mənimsədilir: $\alpha_{t_h}(Q) = 0, \forall t_h \notin T^Q$.

Sorğuların terminlərinin vaciblik dərəcələrinin relevant
sənədlərin seçilməsi prosesində nəzərə alınması məqsədilə
 T^Q çoxluğunun bütün elementləri (yəni sorğunun açar
sözləri) cütləri üçün qeyri-səlis üstünlük münasibətləri
müəyyən edilir:

$$\gamma : T^Q \times T^Q \rightarrow [0,1], \quad (7.28)$$

burada $\gamma(t_u, t_v)$ – vaciblik əmsalı olub t_u termininin t_v ter-
minindən nə dərəcədə daha vacib olduğunu göstərir. $\gamma(t_u, t_v)$
funksiyasının qiymətləri t_u və t_v terminlərinin $\alpha_{t_u}(Q)$ və
 $\alpha_{t_v}(Q)$ vaciblik əmsallarının əsasında hesablanır:

$$\gamma(t_u, t_v) = \begin{cases} 1 - [\alpha_{t_v}(Q) - \alpha_{t_u}(Q)], & \alpha_{t_u}(Q) < \alpha_{t_v}(Q) \text{ olarsa,} \\ 1, & \alpha_{t_u}(Q) > \alpha_{t_v}(Q) \text{ olarsa.} \end{cases} \quad (7.29)$$

Beləliklə, sorğuya relevant olan sənədlərin tapılması məsələsi sorğuda olan bütün axtarış terminlərinə görə ən yaxın sənədlərin seçilməsi prosedurasına çevrilir. Bunun üçün ayrı-ayrı t_v terminləri üzrə D çoxluğunun bütün sənədlərinin üstünlük (daha relevant olma) dərəcələri müəyyən edilir.

Sənədlərin üstünlük dərəcələri onların qeyri-səlis üstünlük münasibətlərinin köməyi ilə təsvir olunur. Bütün axtarış terminləri ilə münasibətlərinə görə D çoxluğunun sənədlərinin bütün cütləri üçün qeyri-səlis üstünlük münasibətləri bu sənədləri təsvir edən terminlərin çəki əmsallarının əsasında hesablanır. Başqa sözlə, $\mu_{t_v}(d_i, d_j)$ daxilolma (üstünlük) funksiyası aşağıdakı kimi müəyyən olunur:

$$\mu_T : D \times D \rightarrow [0,1], \quad (7.30)$$

burada $\mu_{t_v}(d_i, d_j)$ daxilolma funksiyası t_v termininə görə d_i sənədinin d_j sənədindən üstünlük münasibətini müəyyən edir. Onun qiyməti isə d_i sənədinin t_v termininə görə d_j sənədindən üstünlük dərəcəsini, yəni t_v termininin d_i sənədi üçün çəki əmsalının d_j sənədi üçün çəki əmsalından nə qədər böyük olmasını göstərir.

Əgər t_v termini d_j sənədinə nisbətən d_i sənədi üçün daha vacibdirsə, onda fərz edilir ki, d_i sənədi t_v termininə görə d_j sənədindən daha üstündür (yəni daha relevantdir).

$\mu_{t_v}(d_i, d_j)$ daxilolma funksiyasının qiymətləri t_v termininin d_i və d_j sənədləri üçün $\omega_{t_v}(d_i)$ və $\omega_{t_v}(d_j)$ çəki əmsallarının əsasında aşağıdakı şəkildə hesablanır:

$$\mu_{t_v}(d_i, d_j) = \begin{cases} 1 - \left[\frac{\omega_{t_v}(d_j) - \omega_{t_v}(d_i)}{\omega_{t_v}(d_j)} \right], & \omega_{t_v}(d_i) < \omega_{t_v}(d_j) \text{ olarsa,} \\ 1, & \omega_{t_v}(d_i) > \omega_{t_v}(d_j) \text{ olarsa,} \end{cases} \quad (7.31)$$

t_v termini baxımından d_i sənədinin d_j sənədinə nisbətən ciddi üstünlük münasibətini müəyyən edən $\mu_{t_v}^s(d_j, d_i)$ qeyri-səlis funksiyası aşağıdakı düsturla ifadə olunur:

$$\mu_{t_v}^s(d_j, d_i) = \begin{cases} \mu_{t_v}(d_i, d_j) - \mu_{t_v}(d_j, d_i), & \mu_{t_v}(d_i, d_j) > \mu_{t_v}(d_j, d_i) \text{ olarsa,} \\ 0, & \mu_{t_v}(d_i, d_j) < \mu_{t_v}(d_j, d_i) \text{ olarsa} \end{cases} \quad (7.32)$$

D çoxluğunun sənədlərinin bütün cütləri arasındakı bu qeyri-səlis üstünlük münasibətlərini nəzərə almaqla konkret təsbit olunmuş $t_v \in T^Q$ termininə görə daha üstün olan sənədlərin qeyri-səlis alt çoxluğunu müəyyən etmək olar. Həmin sənədlərin bu çoxluğa daxilolma funksiyası aşağıdakı şəkildə verilə bilər:

$$\mu_{t_v}^{pr.}(d_i) = 1 - \max_{d_j \in D} \left\{ \mu_{t_v}^s(d_j, d_i) \right\}, \quad (7.33)$$

burada $\mu_{t_v}^{pr.}(d_i)$ daxilolma funksiyasının qiyməti t_v termini üzrə d_i sənədinin D çoxluğunun digər sənədlərinə nisbətən nə qədər üstün (daha relevant) olmasını göstərir.

Sorğunun ayrı-ayrı terminləri üzrə D çoxluğunun bütün sənədlərinin üstünlük dərəcələri, yəni qeyri-səlis üstün sənədlər alt çoxluğu məlum olduqdan sonra bu alt çoxluğa daxil olan sənədlərin cütləri arasında sorğunun bütün terminlərini nəzərə almaqla qeyri-səlis üstünlük münasibətlərini müəyyən etmək olar. Bu məqsədlə T^Q çoxluğuna daxil olan terminlərin cütləri arasındakı qeyri-səlis üstünlük münasibətləri $\gamma(t_u, t_v)$ istifadə oluna bilər:

$$\mu_{T^Q}(d_i, d_j) = \max_{t_u \in T^Q} \left\{ \min_{t_v \in T^Q} \left\{ \mu_{t_u}^{pr.}(d_i), 1 - \mu_{t_v}^{pr.}(d_j), \gamma(t_u, t_v) \right\} \right\}, \quad (7.34)$$

burada $\mu_{T^Q}(d_i, d_j)$ – D çoxluğunun sənədlərinin cütləri arasında qeyri-səlis üstünlük münasibəti ilə müəyyən olu-

nan alt çoxluğuna daxilolma funksiyasıdır və onun qiyməti sorğunun bütün terminlərinə (yəni bütövlükdə sorğuya) görə d_i sənədinin d_j sənədindən üstünlük dərəcəsini müəyyən edir.

Yuxarıda baxılan yanaşmaya analogi olaraq, D çoxluğunun bütün sənəd cütləri üçün tapılmış qeyri-səlis üstünlük münasibətlərinin köməyi ilə bu çoxluğun qeyri-səlis üstünlüyə malik olan sənədləri alt çoxluğunu, yəni sorğu üçün bütün sənədlərin üstünlük dərəcələrini müəyyən etmək olar:

$$\mu_{TQ}^{pr}(d_i) = 1 - \max_{d_j \in D} \left\{ \mu_{TQ}^s(d_j, d_i) \right\}, \quad (7.35)$$

burada $\mu_{TQ}^s(d_i, d_j)$ funksiyasının qiymətləri aşağıdakı kimi hesablanır:

$$\mu_{TQ}^s(d_j, d_i) = \begin{cases} 0, & \mu_{TQ}(d_j, d_i) \leq \mu_{TQ}(d_i, d_j) \text{ olarsa,} \\ \mu_{TQ}(d_i, d_j) - \mu_{TQ}(d_j, d_i), & \mu_{TQ}(d_j, d_i) > \mu_{TQ}(d_i, d_j). \end{cases} \quad (7.36)$$

Beləliklə $\mu_{TQ}^{pr}(d_i)$ daxilolma funksiyası qeyri-səlis üstünlük münasibətlərinə malik sənədlər alt çoxluğunu xarakterizə edir. Bu funksiya sorğu üçün d_i sənədinin üstünlük dərəcəsini, yəni onun istifadəçinin sorğusuna relevantliq səviyyəsini göstərir.

Təbii ki, $\mu_{TQ}^{pr}(d_i)$ funksiyası $[0,1]$ intervalında istənilən qiyməti ala bilər, lakin istifadəçini yalnız yüksək relevantliq dərəcəsi olan sənədlər maraqlandırır. Bu amilin nəzərə alınması məqsədilə üstünlük, yəni relevantliq dərəcələri üçün $\varepsilon \in [0,1]$ sərhəd qiymətini (adətən, bu qiymət $\varepsilon > 0.5$ qəbul olunur) daxil etmək olar. Onda

$$\mu_{TQ}^{pr}(d_i) \geq \varepsilon \quad (7.37)$$

şərtini ödəyən bütün sənədlər sorğuya daha relevant sənədlər kimi qəbul olunur.

Beləliklə,

$$D^Q = \{d_i; d_i \in D; \mu_{T^Q}^{pr}(d_i) \geq \varepsilon\} \quad (7.38)$$

sənədlər alt çoxluğu sorğuya relevant olan sənədlər çoxluğunu təşkil edir və axtarışın nəticəsi kimi istifadəçilərə verilir. Daxilolma funksiyası

$$\mu(d^*) = \max_{d_i \in D^Q} \{\mu_{T^Q}^{pr}(d_i)\} \quad (7.39)$$

olan d^* sənədi onlar arasında ən relevant sənəd olacaqdır.

İnformasiya axtarışı üçün təklif olunan üsul daha relevant olan, yəni istifadəçinin tələbatını daha dəqiq ödəyən sənədləri tapmağa imkan verir. Başqa sözlə, bu üsul internetin informasiya fəzasının axtarış sistemi tərəfindən indeksləşdirilmiş bütün sənədlərini bir-biri ilə müqayisə edərək onlar arasından daha üstün olanlarını tapır. Bu zaman sənədlər arasındakı münasibətlərlə yanaşı, terminlərlə sənədlər, habelə terminlərin özləri arasındakı münasibətlər də nəzərə alınır.

Burada nəzərə alınmalıdır ki, təklif olunan üsul məhdud sənədlər çoxluğu tətbiqi üçün tətbiqi daha səmərəli nəticələr verir. Belə ki, qapalı (yuxarıdan məhdud olan) sənədlər çoxluğunda cüt-cüt müqayisənin aparılmasını daha asan reallaşdırmaq mümkündür.

7.5. Qeyri-səlis üstünlük münasibətləri əsasında paylanmış informasiya axtarışının təşkili

Yuxarıda qeyd olunduğu kimi, internetdə çoxlu sayda axtarış sistemləri mövcuddur. Bu sistemlər müəyyən üstün və çatışmayan cəhətlərə malik olurlar. Müxtəlif axtarış maşınları informasiya ehtiyatlarını (web-səhifələri, məlu-

mat bazalarını, fayl arxivlərini və s.) özlərinə məxsus şəkildə indeksləşdirirlər [95]. Nəticədə, tərtib olunmuş eyni sorğular üzrə ayrı-ayrı İAS tərəfindən aparılan axtarışlar müxtəlif nəticələr verir. Həmçinin, axtarış nəticəsində ayrı-ayrı axtarış maşınları tərəfindən tapılmış eyni sənədlərin relevantlıq əmsalları fərqli ola bilər. Bundan əlavə, axtarış sistemləri tematik baxımdan istiqamətlənmiş ola bilər ki, bu da öz əksini axtarışın nəticəsində tapa bilər.

Ona görə də müxtəlif axtarış sistemlərinin əldə etdikləri nəticələrin müqayisəli təhlili və onlar tərəfindən eyni sorğuya verilmiş sənədlər arasından daha münasib (relevant) sənədlərin seçilməsi məsələsi böyük aktuallıq kəsb edir. Qeyd olunmalıdır ki, axtarış maşınları tərəfindən tapılıb verilmiş sənədlər çoxluğu informasiya fəzasında olan və axtarış sistemləri tərəfindən əhatə edilən sənədlər çoxluğundan xeyli azlıq təşkil edir. Güman etmək olar ki, axtarış sistemləri tərəfindən tapılıb verilmiş sənədlər çoxluğunun təhlili və emalı paylanmış informasiya axtarışını həyata keçirməyə və axtarışın nəticəsini xeyli yaxşılaşdırmağa imkan verə bilər.

Paylanmış informasiya axtarışının təşkili İAS-ın aşağıdakı əsas xarakteristikalarını yaxşılaşdırmağa imkan verir:

- altsistemlərin paralel yerinə yetirilməsi yolu ilə axtarışın tez həyata keçirilməsi;
- sistemin uzaqda olan komponentlərini cəlb etmədən lokal qərarların müstəqil qəbul edilməsinin mümkünlüyü;
- modulluluq – altsistemlərin işlənilib hazırlanmasının bir-birindən asılı olmaması, sistemin genişləndirilməsinin asanlıığı və s.;
- sıradan çıxmaya davamlılıq – sistemin bir altsisteminin və ya onun qoşulduğu rabitə kanalının sıradan

çıxması digər altsistemlərin, eləcə də bütövlükdə sistemin işinin dayanmasına səbəb olmur;

- şəbəkə və zaman resurslarının qənaət olunması və s.

Paylanmış axtarışın təşkili bir çox alt sistemlərdən (axtarış maşınlarından və robotlarından) ibarət paylanmış İAS-ın – multi-axtarış sisteminin yaradılması zamanı daha böyük əhəmiyyət kəsb edir. Alt sistemlər tərəfindən verilmiş lokal qərarlar təhlil olunur, ümumiləşdirilir və bundan sonra onlar arasında seçim aparılır. Nəticədə, seçilmiş sənədlər multi-axtarış sisteminin nəticəsi kimi qəbul olunur və istifadəçiyə təqdim edilir.

Bu paraqrafda axtarış alt sistemlərinin lokal qərarlarının müqayisəli təhlili və onlar arasında olan qeyri-səlis üstünlük münasibətlərinə görə ümumi qərarın qəbul edilməsi yolu ilə paylanmış informasiya axtarışının təşkili üsuluna baxılır. Başqa sözlə, çoxlu sayda axtarış maşınlarından ibarət olan multi-axtarış sisteminin yaradılması və bu axtarış maşınlarının verdikləri nəticələr əsasında ümumi qərarın qəbul edilməsi məsələsi tədqiq olunur.

Tutaq ki, $S = \{s_k, k = \overline{1, m}\}$ – multi-axtarış sisteminin axtarış maşınları çoxluğu, $D = \{d_i, i = \overline{1, n}\}$ – arasından daha relevant olanların seçilməsi tələb olunan sənədlər çoxluğu, $D_k, k = \overline{1, m}$ – s_k axtarış maşını tərəfindən tapılmış sənədlər alt çoxluğu, Q – istifadəçi sorğusudur.

Qeyd olunmalıdır ki, $D_k, k = \overline{1, m}$ alt çoxluqları aşağıdakı şərtləri ödəyirlər:

1. $D = D_1 \cap D_2 \cap \dots \cap D_m$;
2. $D_i \cup D_j \neq \emptyset, \forall i \neq j$ üçün.

Tutaq ki, s_k axtarış maşını tərəfindən tapılıb verilmiş hər bir d_i sənədi istifadəçinin sorğusunu təyin oblastı $[0,1]$ intervalında olan $\mu_{s_k}(d_i)$ daxilolma funksiyası ilə müəyyən edilən dərəcədə ödəyir. Bu funksiyanın qiymətlərinə əsasən axtarış maşınları tərəfindən verilmiş sənədlər çoxluğunda qeyri-səlis üstünlük münasibətlərini müəyyən etmək olar. Bu münasibət s_k axtarış maşınının nəticələrinə görə d_i sənədinin d_j sənədindən üstünlük dərəcəsini ifadə edən $\psi_{s_k}(d_i, d_j)$ daxilolma funksiyası ilə təsvir olunur. Bu funksiya $[0,1]$ intervalında qiymətlər alır və aşağıdakı düsturla müəyyən edilir:

$$\psi_{s_k}(d_i, d_j) = \begin{cases} 1 - \left| \frac{\mu_{s_k}(d_j) - \mu_{s_k}(d_i)}{\mu_{s_k}(d_j)} \right|, & \mu_{s_k}(d_j) \geq \mu_{s_k}(d_i) \text{ olarsa,} \\ 1, & \text{digər halda} \end{cases} \quad (7.40)$$

burada $k = \overline{1, m}$, $i, j = \overline{1, n}$.

(7.40) düsturunun köməyi ilə bütün axtarış maşınlarının nəticələri əsasında sənədlərin qeyri-səlis üstünlük münasibətləri tapılır. Lakin nəzərə alınmalıdır ki, multi-axtarış sisteminə daxil olan axtarış maşınlarının konkret axtarış istiqamətləri üzrə nəticəliliyi (səriştəliliyi, kompetentliyi), eləcə də coğrafi yerləşməsi, kanalların buraxılış qabiliyyətləri, internetin informasiya fəzasının zonalarının əhatə olunması və s. ilə bağlı xarakteristikaları müəyyən obyektiv səbəblərdən eyni olmur.

Tutaq ki, multi-axtarış sisteminin axtarış maşınlarının səriştəlilik dərəcələri $[0,1]$ intervalında təyin olunan $\gamma(s_k)$ əmsalı ilə ifadə edilir. $\gamma(s_k)$ əmsallarının əsasında axtarış maşınlarının qeyri-səlis üstünlük münasibətləri təyin edilir:

$$\gamma(s_i, s_j) = \begin{cases} 1 - [\gamma(s_j) - \gamma(s_i)], & \gamma(s_j) \geq \gamma(s_i) \text{ olarsa,} \\ 1, & \end{cases} \quad (7.41)$$

burada $\gamma(s_i, s_j) - s_i$ axtarış maşınının s_j axtarış maşınından daha səriştəli olması dərəcəsinə göstərir.

İnternetin informasiya fəzasından istifadəçi sorğusuna relevant olan sənədlərin axtarılması məsələsi axtarış maşınları tərəfindən tapılmış sənədlər çoxluğundan yuxarıda təsvir olunmuş məlumatlara əsasən daha relevant olanların seçilməsi məsələsinə gətirilir.

Əvvəlcə multi-axtarış sisteminin hər bir $s_k \in S$ axtarış maşını üçün qeyri-səlis üstünlük münasibətlərini təşkil edən sənədlər çoxluğunu tapaq. Bunun üçün s_k axtarış maşınlarına görə d_i və d_j sənədləri arasındakı $\psi_{s_k}(d_i, d_j)$ qeyri-səlis üstünlük münasibətlərini istifadə etmək olar. Belə ki, üstünlük təşkil edən sənədlər çoxluğu aşağıdakı daxilolma funksiyası ilə təyin olunur:

$$\psi_{s_k}^n(d_i) = 1 - \sup_{d_j \in D} [\psi_{s_k}(d_i, d_j) - \psi_{s_k}(d_j, d_i)]. \quad (7.42)$$

Aydındır ki, $\psi_{s_k}^n(d_i)$ daxilolma funksiyası ən böyük qiymət alan d_i sənədləri s_k axtarış maşınının fərdi qərarını (əldə etdiyi nəticəni) təşkil edir. Digər axtarış maşınlarının əldə etdikləri nəticələri nəzərə almaq üçün sənədlərin ümumiləşmiş qeyri-səlis üstünlük münasibətlərini istifadə etmək olar. Bu münasibətlər axtarış maşınlarının $\gamma(s_i, s_j)$ üstünlük əmsallarını istifadə etməklə aşağıdakı kimi müəyyən olunur:

$$\eta(d_i, d_j) = \sup_{s_k, s_l \in S} \min \{ \psi_{s_k}^n(d_i), \psi_{s_l}^n(d_j), \gamma(s_k, s_l) \}. \quad (7.43)$$

(7.43) qeyri-səlis üstünlük münasibəti mövzu sahəsində sistemin axtarış maşınlarının səriştəliliyi haqqında məlumatlar nəzərə alınmaqla $\psi_{s_k}(d_i, d_j)$ qeyri-səlis münasibətləri sinfinin vahid yekun qeyri-səlis münasibətə "yığılması" deməkdir. Beləliklə, sənədlərin bir neçə üstünlük münasibətləri ilə seçilməsi məsələsi sənədlərin yeganə üstünlük münasibəti ilə seçilməsi məsələsinə gətirilir. Bu məsələnin həlli üçün sənədlər çoxluğunda $\eta(d_i, d_j)$ ümumiləşmiş qeyri-səlis üstünlük münasibətləri əsasında müvafiq sənədlər çoxluğu müəyyən edilir:

$$\tilde{\eta}^n(d_i) = 1 - \sup_{d_j \in D} [\eta(d_i, d_j) - \eta(d_j, d_i)] . \quad (7.44)$$

Nəhayət, bu çoxluqdan üstün olan sənədlərin qeyri-səlis alt çoxluğu tapılır:

$$\eta^n(d_i) = \min[\tilde{\eta}^n(d_i), \eta(d_i, d_j)] . \quad (7.45)$$

Bu çoxluğa daxil olan sənədlər axtarış sistemi tərəfindən verilmiş daha relevant sənədləri özündə birləşdirir. Aşağıdakı düsturla onlar arasından ən relevant olan bir sənədi müəyyən etmək mümkündür:

$$\eta(d^*) = \max_{d_i \in D} [\eta^n(d_i)] \quad (7.46)$$

Beləliklə, paylanmış informasiya axtarışı üçün təklif olunan üsul informasiya fəzasının bütün sənədləri arasından ayrı-ayrı axtarış sistemləri tərəfindən tapılıb verilmiş sənədlər çoxluğundan müqayisəli təhlil yolu ilə daha üstün və ən relevant sənədlərin tapılmasını təmin edir.

VIII FƏSİL

İNTERNETDƏ İNFORMASIYA AXTARIŞI

**İnternetdə informasiyanın axtarılması və əldə edilməsi
üsulları**

İnternetin informasiya-axtarış sistemləri

İnternetin informasiya-axtarış sistemlərinin təsnifatı
(Tematik kataloqlar, axtarış maşınları, meta-axtarış sistemləri)

**İnternetin informasiya ehtiyatlarının
indeksləşdirilməsi və təqdim olunması mexanizmləri**

Axtarış spamları və onlara qarşı mübarizə

8.1. İnternetdə informasiyanın axtarılması və əldə edilməsi üsulları

İnternetin informasiya ehtiyatlarının həcmnin və müxtəlifliyinin kəskin artması səbəbindən bu nəhəng fəzada lazım olan informasiyanın axtarılması və tapılması çox ciddi problemə çevrilmişdir. Aydınır ki, gündəlik praktiki fəaliyyətində müəyyən informasiyaya tələbatı yaranan hər bir insan bu tələbatını ödəmək üçün informasiya mənbələrinə, o cümlədən internetə müraciət edir, müvafiq axtarış sorgusunu formalaşdırır və informasiya xidmətlərinə təqdim etməklə özünə lazım olan informasiyanı axtarır.

İnternetdə informasiyanın axtarılması və əldə edilməsi bir neçə yolla həyata keçirilə bilər. Bu üsullar axtarışın effektivliyini və keyfiyyətini, eləcə də əldə olunub götürülən informasiyanın növünə görə əhəmiyyətli dərəcədə fərqlənirlər. İnformasiya axtaran şəxsin məqsəd və vəzifələrindən asılı olaraq, internetdə informasiyanın axtarılması və əldə olunması üçün əsas üç üsuldən (yanaşmadan) istifadə olunur:

- URL-ünvanın daxil edilməsi;
- hiperəlaqələr üzrə hərəkət (naviqasiya);
- informasiya-axtarış sistemlərinin istifadəsi.

Birinci yanaşma internetdə lazım olan informasiyanı əldə etmək üçün onun yerləşdiyi saytın URL-ünvanını daxil etməklə informasiya ehtiyatına birbaşa qoşulmanı və ya müraciəti nəzərdə tutur. Bu üsul informasiya axtarışının ən sadə və tez başa gələn üsuludur, lakin informasiya ehtiyatının və ya onun yerləşdiyi saytın ünvanı dəqiq məlum olduqda bu üsuldən istifadə etmək mümkündür. Belə ki, bu halda istifadəçi lazım olan məlumatların harada yerləş-

diyini, saxlanma formasını, girişə (istifadəyə) icazə və ya qeydiyyat (əgər tələb olunursa) üçün hara müraciət etmək lazım olduğunu, sistemin iş prinsiplərini, konkret informasiyaya daxilolma və onun əldə edilməsi üsullarını dəqiq bilməlidir. Yalnız bu halda o, kompüter arxasında işləmək və şəbəkə vasitəsilə informasiya sisteminə qoşulmaq yolu ilə lazımi informasiya ehtiyatlarına birbaşa daxil olur və zəruri informasiyanı əldə edir.

İkinci yanaşma hər hansı internet saytında mövcud hiperistinadlardan istifadə etməklə lazım olan zəruri informasiya ehtiyatına keçilməsini nəzərdə tutur. Web-fəzada bütün saytlar bir-biri ilə məzmunca hiperistinadlarla bağlı olduqlarına görə informasiya axtarışı web-səhifələrin bu əlaqələr vasitəsilə ardıcıl baxılması yolu ilə həyata keçirilə bilər. Əl ilə axtarış prinsipinə əsaslanan bu üsul milyonlarla saytlar üzrə adi baxma yolu ilə lazım olan informasiyanın axtarılması əlverişli deyildir, lakin axtarışın son mərhələsində, informasiya-axtarış sistemləri tərəfindən verilmiş nəticələrin baxılması, konkret mövzu sahəsinə aid saytda lazım olan məlumatın axtarılması, daha dərin məzmunun təhlilinə zərurətin yaranması zamanı çox faydalı olur. Aydın ki, informasiya ehtiyatının ünvanı istifadəçiyə bəlli olmadıqda bu üsuldən istifadə rahatdır. Belə yanaşmada axtarışa başlamanın ilkin nöqtəsi kimi məşhur web-portallardan istifadə edilə bilər.

Üçüncü yanaşma internetdə sənədlərin axtarışının aparılması üçün əsas və ən çox istifadə olunan üsullardan biridir. İnternetin informasiya-axtarış sistemləri internetin informasiya ehtiyatları haqqında əvvəlcədən, xüsusi formada hazırlanmış məlumat bazası, bu bazanı yaradan və daim yeniləşdirən proqram təminatı, axtarış sorğularının daxil edilməsi və nəticələrin istifadəçiyə təqdim olunması

üçün interfeys, xüsusi axtarış mexanizmləri (açar sözlərə görə axtarış və ya mövzü kataloqları üzrə axtarış) olan proqram kompleksidir. Bu sistemlər axtarışın nəticəsi olaraq istifadəçiyə onların sorğusuna relevant olan sənədlərin siyahısını təqdim edir. Həmin siyahıda təqdim olunan sənədlərin müəyyən hissəsi (bəzən əksəriyyəti) əslində istifadəçiyə lazım olmur. Ümumiyyətlə, nəticədə verilən sənədlərin diqqətlə araşdırılması lazım gəlir. İnternetin informasiya-axtarış sistemləri növbəti bölmələrdə ətraflı baxılacaqdır.

İnternetdə lazımı informasiyanın axtarılması və əldə olunması üçün yuxarıda göstərilən üsullardan əlavə praktikada tez-tez istifadə olunan daha iki yanaşmanı qeyd etmək olar:

- yeni informasiya ehtiyatlarının təhlili;
- xüsusi proqramların tətbiqi yolu ilə zəruri informasiya ehtiyatının axtarılması.

İnternetdə yeni yaranmış informasiya ehtiyatlarının baxılması və təhlili təkrar axtarışların aparılması, ən yeni informasiyanın tapılması və ya haqqında informasiya lazım olan obyektin inkişaf tendensiyasının izlənməsi zamanı çox əlverişli olur. Bu zaman həmin mövzu sahəsi ilə bağlı bütün informasiya ehtiyatlarının baxılmasına ehtiyac olmur, istifadəçini yalnız internetdə yeni dərc olunmuş məlumatlar maraqlandırır.

İnformasiyanın ilkin axtarışının aparılması üçün xüsusi proqramların tətbiqi effektiv və avtomatlaşdırılmış üsul hesab edilir. Yuxarıda ikinci yanaşmada təklif olunan əl ilə həyata keçirilən hiperistinadlar üzrə hərəkət burada avtomatlaşdırılır. Konkret mövzu sahəsi üzrə verilmiş sözlər (terminlər) üzərində qoyulmuş hiperistinadların göstərdiyi web-səhifələrin baxılması spyayder adlanan xüsusi proqramlar vasi-

təsilə avtomatlaşdırılmış şəkildə həyata keçirilir. Belə proqramlar üçüncü yanaşmada təqdim olunan informasiya-axtarış sistemlərinin robot proqramlarının xüsusi formasıdır. Bu yanaşmaya, həmçinin, dar mövzu sahəsində informasiya axtarışını həyata keçirən xüsusi proqramlar da aid edilir. Təbii ki, bu yanaşma nəticəsində təqdim olunan sənədlərin də istifadəçi tərəfindən dəqiq emalı tələb olunur.

Ümumiyyətlə, internetdə informasiya axtarışı dedikdə mövcud informasiya-axtarış sistemləri, eləcə də hər hansı saytın və ya informasiya sisteminin axtarış mexanizmləri vasitəsilə lazım olan məlumatın axtarılıb tapılması nəzərdə tutulur. Bu halda istifadəçi ona hansı informasiyanın lazım olduğunu bilir, lakin bu informasiyanın harada, hansı internet saytında və ya informasiya sisteminə yerləşdiyini bilmir. İnternetdə axtarış aparmaq üçün istifadəçi zəruri axtarış sorğusunu formalaşdıraraq müvafiq informasiya-axtarış sistemə və ya axtarış mexanizmi olan informasiya sistemə ötürür. İstifadəçi sorğusunun cavabında axtarış sistemi tərəfindən tapılmış informasiya ehtiyatlarının siyahısını (adları və ya ünvanları göstərilməklə) alır.

Əvvəlki URL-ünvanın daxil edilməsi yolu ilə lazım olan saytın açılması, hiperistinadlar üzrə hərəkət etməklə zəruri informasiya ehtiyatına keçid və yeni informasiya ehtiyatlarının təhlili kimi yanaşmaları isə sadəcə internetdə naviqasiya adlandırmaq olar.

8.2. İnternetin informasiya-axtarış sistemləri

İnternetin informasiya-axtarış sistemi – axtarış mexanizmləri ilə təchiz olunmuş və yalnız verilmiş sistem (sayt) daxilində deyil, eləcə də şəbəkə vasitəsilə internet

informasiya fəzasında mövcud olan bütün saytlar, sistemlər, məlumat bazaları və digər informasiya ehtiyatlarında axtarışının aparılması üçün nəzərdə tutulmuş informasiya sistemidir. Sadəlik naminə gələcəkdə informasiya-axtarış sistemi anlayışında istifadə edəcəyik.

Ümumiyyətlə, İAS-a internetin informasiya fəzasında informasiya xidmətlərinin istifadəçiləri üçün bələdçi kimi baxmaq olar. Adətən, istifadəçilər özlərinə hansı informasiyanın lazım olduğunu yaxşı bilir, lakin bu informasiyanın şəbəkədə mövcud olmasına baxmayaraq, onu harada və necə tapmağın mümkün olduğunu bilmirlər. Məhz belə problemin həlli və istifadəçilərin informasiyaya olan tələbatının ödənilməsində onlara yardım göstərən bələdçi rolunu oynayan vasitəçi proqram sistemlərindən, yəni İAS-dan istifadə olunur. İAS müxtəlif yollarla (informasiya ehtiyatlarının tematik kataloqlarını və ya avtomatik indekslərini yaratmaq, onların sahiblərindən, o cümlədən serverlərdən zəruri informasiyanı sifariş etmək yolu ilə və s.) internet şəbəkəsində olan informasiya ehtiyatlarının kontenti, yəni məzmunu haqqında "bilikləri" öz bazasına toplayırlar.

İnternetin müasir informasiya xidmətlərinin, o cümlədən İAS-ın əsasında ənənəvi İAS-ın yaradılması və istifadəsi sahəsindəki təcrübə durur.

İnternetin informasiya ehtiyatlarının xüsusiyyətlərini nəzərə alaraq, yaradılan informasiya axtarışı vasitələrinə, o cümlədən İAS-a qoyulan əsas tələblər aşağıdakılardır:

- İnternetin informasiya ehtiyatlarının maksimal dərəcədə dolğun əhatə olunması;
- İnternetdən alınmış (götürülmüş) informasiyanın doğruluğunun təmin edilməsi və ziddiyyətliliyinin aradan qaldırılması;

- informasiya ehtiyatlarının avtomatik indeksləşdirilməsi və tematik kataloqlarının yaradılması üçün intellektual üsulların reallaşdırılması;
- istifadəçiyə vaxtında və dolğun cavab vermək üçün onun sorğusuna uyğun olaraq axtarış sistemi tərəfindən yüksəksürətli axtarışın aparılmasının təmin edilməsi;
- axtarışın nəticəsinin dəqiqliyinin və pertinentliyinin, yəni alınmış nəticələrin istifadəçinin tələbatına tamamilə uyğun olmasının təmin edilməsi;
- istifadəçi ilə əks əlaqənin qurulması imkanlarının reallaşdırılması;
- təbii dilə yaxın olan istifadəçi interfeysinin reallaşdırılması.

Qeyd olunan tələblərin ödənilməsi üçün İAS tərəfindən aşağıdakı əsas məsələlərin həlli təmin edilməlidir:

- sənədlərin (informasiya ehtiyatlarının) indeksləşdirilməsi, onların axtarış obrazlarının yaradılması və İAS-ın indekslər bazasında təqdim olunması, saxlanması və yenilənməsi;
- terminlərin, sinonimlərin, tezaurusların və assosiativ sözlərin tələb olunan rubrikatorlarının, lüğətlərinin, so-raqçalarının, eləcə də məlumat bazalarının yaradılması;
- sənədlərin tematikalarının müəyyənləşdirilməsi, onların tematik kataloqlarının yaradılması, eləcə də müəyyən olunmuş rubrikatora görə sənədlərin müvafiq tematik kataloqlar üzrə paylanması;
- strukturlaşdırılmış indekslər bazasının və meta-verilənlər bazalarının yaradılması;
- istifadəçi ilə axtarış sistemi arasında intellektual interfeysin qurulması;

- istifadəçilərin sorğularının emalı və indeksləşdirilməsi, axtarış obrazlarının yaradılması;
- istifadəçinin təbii dildə mətn şəklində tərtib etdiyi axtarış sorğusuna əsasən onun informasiya tələbatının müəyyən edilməsi;
- istifadəçi sorğusuna relevant olan sənədlərin indekslər bazasında və tematik kataloqda axtarışının təşkili;
- tapılan sənədlərin istifadəçi sorğusuna relevantlıq dərəcələrinin müəyyənləşdirilməsi və nəticələrin nizamlanmış şəkildə istifadəçiyə təqdim olunması;
- istifadəçinin sistemlə qarşılıqlı fəaliyyətinin təşkili və effektiv istifadəçi interfeysinin yaradılması;
- istifadəçi ilə əks əlaqə mexanizmlərinin reallaşdırılması.

İnternetdə informasiya ehtiyatlarının həcminin astronomik sürətlə artması, onlar üzrə axtarışı həyata keçirən İAS-a müraciətlərin sayının və intensivliyinin kəskin yüksəlməsi və digər səbəblərdən aşağıdakı məsələlər də bu gün böyük aktuallıq kəsb edir:

- axtarış vasitələrinin, indekslər bazalarının və tematik kataloqların internetin qovşaqları və serverləri üzrə paylanması, yəni paylanmış informasiya axtarışının təşkili;
- axtarış agentlərinin yaradılması, multiagent axtarış sistemlərinin yaradılması;
- sənədlərin tematikasının avtomatik müəyyənləşdirilməsi və onların tematik kataloqlarının avtomatik yaradılması;
- təbii dilə yaxın intellektual istifadəçi interfeysinin yaradılması;
- informasiya axtarışı vasitələrinin işlənilib hazırlanması və reallaşdırılması üçün qeyri-səlis çoxluqlar, süni intellekt və neyron şəbəkələr nəzəriyyələrinin imkanlarından istifadə edilməsi.

8.3. İnternetin informasiya-axtarış sistemlərinin təsnifatı

Hazırda internetdə çoxlu sayda informasiya-axtarış sistemləri fəaliyyət göstərir. Onların əsas məqsədi istifadəçilərə zəngin servis, lazım olan məlumatlara sürətli və rahat girişi təmin edən informasiya xidmətinin təşkilindən ibarətdir. Belə ki, istifadəçi internetin informasiya sistemlərinə (saytlara) bilavasitə müraciət etmədən İAS-ın köməyi ilə həmin saytlarda saxlanılan informasiya ehtiyatları haqqında ətraflı məlumat əldə etmək, o cümlədən onların URL-ünvanlarını, formatını, dilini, tematikasını və s. müəyyənləşdirmək imkanına malik olur.

Ümumiyyətlə, İAS-ı bir sıra kriterilərə görə təsnif etmək olar. Əvvəla, web-mühitdə reallaşdırma zamanı istifadə olunan yanaşmalara və sənədlərin axtarış obrazlarının yaradılması prinsiplərinə görə informasiya-axtarış sistemlərini aşağıdakı növlərə bölmək olar:

- tematik (mövzu) kataloqlar;
- axtarış maşınları – avtomatik indekslər;
- meta-axtarış sistemləri.

Qeyd edilməlidir ki, bu bölgü şərtidir. Belə ki, bütün tematik kataloqlar həm də açar sözlərə görə axtarış funksiyalarını reallaşdırırlar. Axtarış maşınları isə bir çox hallarda kataloq təsnifatlarını özlərində birləşdirirlər.

İnformasiya ehtiyatlarının təsvir olunması üçün terminlərin seçilməsi üsullarına görə axtarış sistemləri iki yerə bölünür:

- terminlərin aid edilməsi üsulundan istifadə edən İAS;
- terminlərin sənədin məzmunundan seçilməsi üsulundan istifadə edən İAS.

Birinci qrupa daxil olan İAS-da terminlərin seçilməsi üçün istifadə edilən üsula əsasən sənədlərə aid edilən və in-

dekslər bazasına yazılan terminlər bu sənədin məzmununda rast gəlinməyə də bilər, lakin mənaca onun məzmununu tam dəqiqliyi ilə əks etdirməlidir. Həmin terminlər xüsusi olaraq hazırlanmış müxtəlif lüğətlərdən, tezauruslardan, soraqçalardan və s. seçilir. *İkinci qrupa* daxil olan İAS-da terminlər xüsusi indeksləşdirmə üsullarının köməyi ilə sənədlərin məzmunlarından seçilir və indekslər bazasına əlavə edilir.

İndeksləşdirmə proseduralarının reallaşdırılması üsullarına görə də İAS-ı iki kateqoriyaya ayırırlar:

- əl ilə indeksləşdirmə üsulları reallaşdırılan İAS;
- avtomatik indeksləşdirmə üsulları reallaşdırılan İAS.

Birinci növ axtarış sistemlərində informasiya ehtiyatlarının indeksləşdirilməsi prosesi mövzu sahəsini yaxşı bilən mütəxəssislər tərəfindən əl ilə, ikinci növ sistemlərdə isə xüsusi proqram vasitələrinin, o cümlədən robot proqramlarının köməyi ilə həyata keçirilir.

İsifadə olunan informasiya-axtarış dillərinə görə İAS iki yerə bölünür:

- sadə dialoqu olan İAS – açar sözlərə görə və ya tematik kataloq üzrə sadə axtarış aparmağa imkan verir;
- inkişaf etmiş dialoqu olan İAS – məntiqi əməliyyatların köməyi ilə mürəkkəb sorğuları tərtib etmək və onların əsasında axtarış həyata keçirmək imkanına malik olurlar.

Nəticələrin istifadəçiyə təqdim edilməsi üsullarına görə İAS-ı iki yerə bölmək olar:

- sənədlərin relevantlıq dərəcələrinin təhlili aparılmadan tərtib olunan sadə siyahı təqdim edən İAS;
- sorğuya relevantlıq dərəcələrinə görə tapılmış sənədlərin düzülmüş – nizamlanmış (sistemləşdirilmiş) siyahısını təqdim edən İAS.

İstifadəçi ilə əlaqənin təşkili üsullarına görə də İAS iki qrupa bölünür:

- istifadəçi ilə əks əlaqəsi olmayan İAS;
- istifadəçi ilə əks əlaqə reallaşdırılan İAS.

8.3.1. Tematik (mövzu) kataloqlar

Tematik kataloqlar İnternetin informasiya ehtiyatlarının mövzular üzrə sistemləşdirilmiş iyerarxik sistemidir. Tematik kataloqlara bəzən *web-kataloqlar*, *axtarış kataloqları* və ya sadəcə *kataloq* adlandırırlar.

Tematik kataloqların məlumat bazaları informasiya ehtiyatları haqqında zəruri məlumatları özündə saxlayır. Belə ki, internetin informasiya ehtiyatları haqqında məlumatlar toplanır və qabaqcadan müəyyən edilmiş mövzulara uyğun olaraq qruplara – tematik kataloqlara bölünür. Tematik kataloqlar da öz növbəsində ona daxil olan alt mövzulara uyğun olaraq alt tematik kataloqlara bölünür, bu alt kataloqlar da daha dar istiqamətli mövzular üzrə alt kataloqlara bölünür və s.

Beləliklə, informasiya ehtiyatları bir-birinə daxil olan iyerarxik mövzular sistemi üzrə təsnif olunur. İnformasiya ehtiyatlarının axtarışı onların mövzularının konkretləşdirilməsi və lokallaşdırılması (yerinin müəyyən edilməsi) yolu ilə həyata keçirilir. Nəticədə, tematik kataloqun iyerarxik strukturu üzrə ən yuxarı səviyyədə başlayaraq ən aşağıdakı səviyyəyə doğru ardıcıl keçmə yolu ilə bütün informasiya ehtiyatlarına giriş təmin edilir.

Tematik kataloqlar klassik kitabxanaların ənənəvi kataloqlarına analogiya təşkil edir və sistemin operatoru və ya administratoru tərəfindən əllə və ya xüsusi proqramların

köməyi ilə avtomatik yaradılır. Kataloqa daxil edilən hər bir internet resurs üzrə qısa annotasiya hazırlanır və mövzular üzrə nizamlanır. Tematik kataloqların yaradılması tərtibatçıdan mənbələrin mövzusunun yaxşı bilməsini və İAS-da istifadə olunan inkişaf etmiş informasiya-axtarış dilinin olmasını tələb edir.

İnsan fəaliyyəti sahəsinə aşağıdakı mövzu sahələri aid edilə bilər: elm, təhsil, texnika, incəsənət, iqtisadiyyat, tibb, siyasət, idman, turizm, istirahət, avtomobil, moda, alış-veriş və s. Qeyd edilməlidir ki, kitabxana sahəsindən fərqli olaraq, internetin tematik kataloqlarında mövzu sahələri üçün vahid standart mövcud deyildir. Belə ki, kitabxana sistemlərində UOT (universal onluq təsnifatlandırıcı), KBT (kitabxana biblioqrafik təsnifatlandırıcı), DOT (Dyui onluq təsnifatlandırıcı) və s. standartlardan istifadə olunur. Bir sıra kitabxanalarda isə spesifik təsnifat sistemləri reallaşdırılır.

İnternet şəbəkəsində vəziyyət daha açıqdır. Belə ki, tematik kataloqda seçilən mövzu sahələri və ya məzmunu eyni olan mövzuların adları saytın sahibindən (onun arzusundan) asılı olaraq tamamilə fərqli ola bilər.

Hazırda tematik kataloqların yaradılması işi, əsasən, insanlar, yəni mütəxəssislər tərəfindən əl ilə yerinə yetirilsə də, son zamanlar tematik kataloqların avtomatik yaradılması üsullarının işlənilib hazırlanması istiqamətində elmi tədqiqatlar aparılır və ümidverici nəticələr əldə olunur.

8.3.2. İnternetin axtarış maşınları

Axtarış maşınları – internetdə avtomatik indekslər prinsipi əsasında fəaliyyət göstərən axtarış sistemləridir. İnternetin axtarış maşınları xüsusi proqramlarının köməyi ilə web-fəzada olan informasiya ehtiyatlarını indeksləş-

dirir, məzmunlarını təhlil edər, onların axtarış obrazlarını yaradır və axtarış sisteminin indekslər bazasına daxil edir. Ona görə də onlara bəzən *avtomatik indekslər* deyirlər.

Avtomatik indekslər bazası xüsusi robot (pauk, spider və s.) proqramlarının köməyi ilə internetdə olan ilkin sənədlərin mətnlərindən, başlıqlarından, annotasiyalarından, referatlarından və digər mətn fraqmentlərindən onların məzmunlarını əks etdirən indekslərin (yəni terminlərin, açar sözlərin) seçilməsi yolu ilə yaradılır. Robot proqramı web-fəzada olan informasiya ehtiyatlarını indeksləşdirir, yəni onların məzmunlarını təhlil edərək, ən vacib terminləri və ya açar sözləri seçib ayırır və axtarış sisteminin indekslər bazasına daxil edir. İndeksləşdirilmiş sənədlərin zəruri rekvizitləri (adı, yeri, ünvanı və s.), habelə həmin termin və açar sözlərin bu sənədlər üçün vaciblik dərəcələrini göstərən əmsallar bu terminlər və açar sözlərlə birlikdə indekslər bazasına əlavə edilir.

İnternetin axtarış maşınları informasiya axtarışını reallaşdırmaq üçün, bir qayda olaraq, aşağıda göstərilən *beş əsas məsələnin* həllini təmin edirlər. Bu məsələlərdən bəziləri axtarış bazasının formalaşdırılması üçün əvvəlcədən yerinə yetirilir, bəziləri isə axtarış prosesində həyat keçirilir.

Birinci məsələ informasiya ehtiyatlarının internetdə yerləşdiyi yerlərin müəyyən edilməsindən və bu ehtiyatların axtarış serverinə yüklənməsindən ibarətdir. İnternetin informasiya ehtiyatları və onların yerləşdikləri yerlər barədə əvvəlki fəsildə danışılmışdır. İnteraksiyanın axtarış serverinə yüklənməsi dedikdə isə informasiya ehtiyatlarından (mətnlərdən, şəkillərdən, audio-video informasiyadan və s.) ibarət olan faylların (onların fraqmentlərinin) indeksləşdirmə üçün serverin müvəqqəti bazasına götürülməsi nəzərdə tutulur.

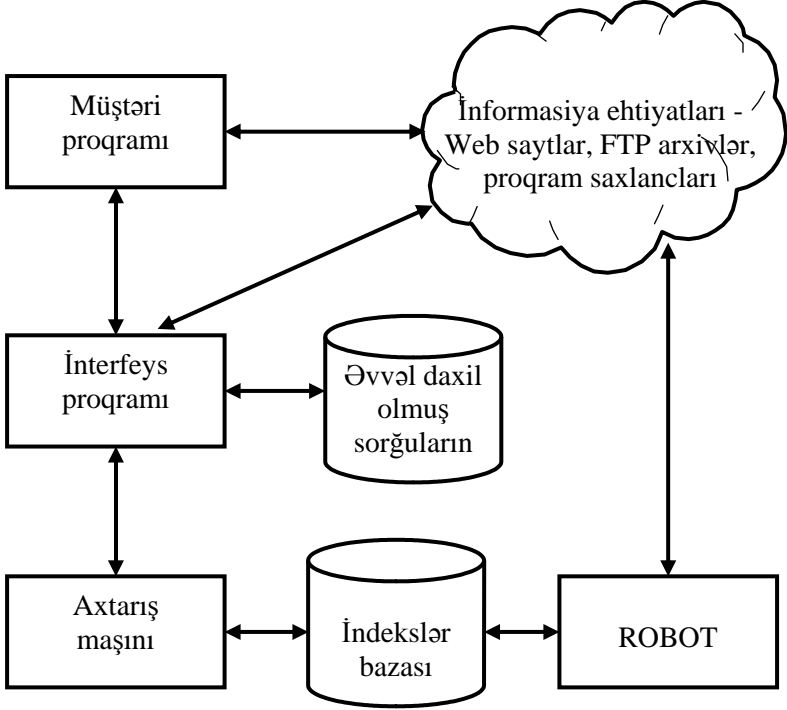
İkinci məsələ – sənədlərin bilavasitə indeksləşdirilməsi, yəni sənədlərin məzmununu dəqiq əks etdirmək üçün onların mətnindən terminlərin və açar sözlərin seçilib çıxarılması, indekslər bazasının yaradılması məsələsidir. Yuxarıda qeyd olunduğu kimi, bu proses indeksləşdirmə adlanır və onun məqsədi sənədlərin axtarış obrazlarının yaradılmasıdır. Təbii ki, internetdə olan informasiya ehtiyatlarının hamısının axtarış serverinin bazasına toplanması və saxlanması mümkün deyil və mənasızdır. Ona görə də İAS-ın bazalarında sənədlərin özləri deyil, onların axtarış obrazları saxlanılır.

Üçüncü məsələ İAS-da istifadəçi interfeysinin, yəni istifadəçi ilə dialoqun təşkilindən ibarətdir. İstifadəçi interfeysi İAS-ın imkanlarını tam əks etdirməli, istifadəçilərə öz tələbatlarını sadə və mürəkkəb sorğular şəklində rahat tərtib etmək, zəruri hallarda onu dəyişdirmək, axtarış sahəsini genişləndirmək və ya daraltmaq imkanları verməli, tapılıb verilmiş sənədlərin sorğuya yaxınlıq dərəcələrinin dəqiqləşdirilməsi yolu ilə sorğunu daha yaxşı formalaşdırmaq və sənədlərin axtarış obrazlarını təkmilləşdirmək üçün istifadəçi ilə əks əlaqəni təmin etməlidir.

İstifadəçinin sorğusuna daha relevant sənədlərin axtarışının bilavasitə indekslər bazasında həyata keçirilməsi İAS-ın qarşısında duran *dördüncü məsələdir*. İstifadəçi tərəfindən axtarış sorğusu formalaşdırılaraq axtarış sisteminə ötürüldükdən sonra axtarış maşını qabaqcadan yaradılmış indekslər bazasından sorğunun tələblərinə cavab verən sənədlər haqqında məlumatları axtarıb tapır və onları istifadəçiyə təqdim edir.

Nəhayət, axtarış maşını tərəfindən tapılmış relevant sənədlərin xüsusi şəkildə sistemləşdirilməsi və nizamlanması, daha relevant sənəddən nisbətən az relevant sənədlərə doğru sıra ilə düzülmiş siyahı şəklində istifadəçiyə verilməsi İAS-ın yerinə yetirdiyi *beşinci məsələni* təşkil edir.

Yerinə yetirdikləri funksiyalara müvafiq olaraq internetin axtarış maşınları tərkibində aşağıdakı komponentləri birləşdirir (şək.8.1):



Şək.8.1. internetin axtarış maşınının ümumi arxitekturası

- **robot proqramı** – internetin informasiya ehtiyatlarını indeksləşdirərək onların obrazlarını yaradan proqramdır;
- **indekslər basası** – internetin informasiya ehtiyatlarının obrazlarını (indekslər çoxluğunu) özündə saxlayan məlumat bazasıdır;

- *axtarış maşını* – istifadəçilərin sorğularına əsasən axtarış sisteminin indekslər bazasında zəruri informasiyanın axtarışını həyata keçirən proqram sistemidir;
- *interfeys proqramı* – lazım olan informasiyanın axtarılması üçün istifadəçilərə informasiya-axtarış dilinin və CGI-skriptlərin köməyi ilə öz sorğularını formalaşdırmaq və axtarış sisteminə ötürmək, eləcə də axtarışın nəticəsini istifadəçiyə təqdim etmək üçün nəzərdə tutulur;
- *müştəri proqramı* – internet vasitəsilə istifadəçilərin İAS-a, daha doğrusu onların serverlərinə qoşulmasına imkan verən proqram təminatıdır. Bu proqram istifadəçilərin kompüterinə yüklənir. Web-səhifələrin baxılması (brauzer) proqramları (Internet Explorer, Netscape Communicator, Opera, Google Chrome və s.) adlanan bu proqramlar İAS-da istifadəçilər üçün nəzərdə tutulmuş interfeys pəncərəsinin istifadəçinin ekranına çıxarılması, web-səhifələrin açılması, bağlanması, köçürülməsi və digər funksiyaları həyata keçirilir.

Göründüyü kimi, internetin informasiya-axtarış sistemləri proqram təminatı və məlumat bazaları kompleksindən ibarətdir. İAS-ın hər bir alt sistemi müstəqil blok olub bir-biri ilə sıx qarşılıqlı əlaqədə işləyir.

8.3.3. Meta-axtarış sistemləri

Məlum olduğu kimi, ayrı-ayrı axtarış maşınları tərəfindən internetin informasiya fəzasının eyni səviyyədə əhatə olunması, eləcə də onların formalaşdırdıqları indekslər bazalarının kəmiyyət və keyfiyyət baxımından fərqli olması

səbəbindən onların eyni sorğuya verdikləri cavablar da eyni olmur. Belə ki, ayrı-ayrı axtarış maşınları tərəfindən tapılıb verilmiş eyni sənədlərin relevantlıq dərəcələri müxtəlif olur. Bununla yanaşı, axtarış maşınları tematik yönümlü ola bilər ki, bu da axtarışın nəticəsinə təsir edə bilər. Ona görə də müxtəlif axtarış maşınlarının eyni sorğu üzrə apardığı axtarış nəticəsində tapdıqları sənədlər çoxluqlarının müqayisəli təhlili və onlar arasından daha münasib (relevant) sənədlərin seçilməsi məsələsi böyük aktuallıq kəsb edir.

Digər təfərdən kifayət qədər nadir məlumatların tapılması məqsədilə bir sıra axtarış sistemlərinə müraciət etmək zərurəti yaranır. Aydın ki, ayrı-ayrı axtarış sistemlərində internet saytların əhatə olunması, indeksləşdirilməsi səviyyələri, eləcə də axtarış sorğularının tərtib edilməsi qaydaları bir-birindən fərqlənir. Ona görə də istifadəçilər həmin sistemlərə növbə ilə müraciət etmək, sorğuların tərtib olunması zamanı onlar xüsusiyyətlərini nəzərə almaq kimi çətinliklərlə üzləşirlər.

Məhz bu səbəbdən internetdə axtarış maşınlarından və tematik kataloqlardan prinsipial olaraq fərqlənən meta-axtarış sistemlərindən istifadə olunur.

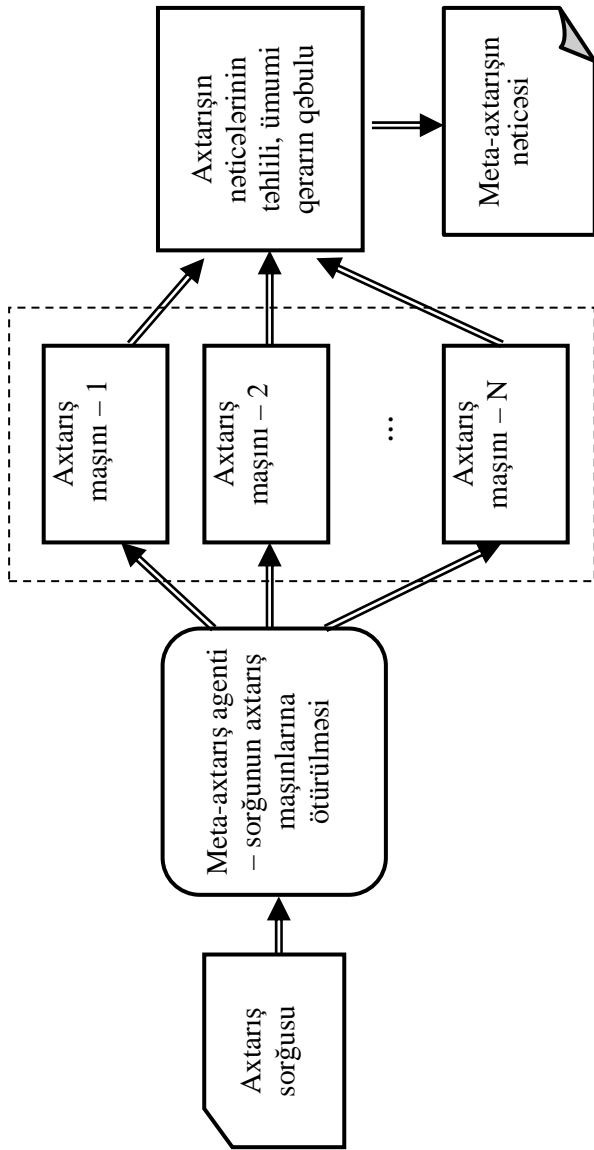
Meta-axtarış sistemləri – internet üçün reallaşdırılan, özlərinin məxsusi indekslər bazası olmaya xüsusi növ axtarış sistemləridir. Onlar istifadəçi ilə interfeysi reallaşdırır, daxil olan axtarış sorğusunu bir neçə axtarış serverinə göndərir, həmin serverlərdən aldığı cavabları xüsusi şəkildə təhlil edərək nəticəni istifadəçiyə təqdim edir.

Meta-axtarış sistemləri faktiki olaraq paylanmış informasiya axtarışını həyata keçirir. Onlar axtarış üçün zəruri olan aşağıdakı əsas xarakteristikaları yaxşılaşdırmağa imkan verir:

- *paralellik* – axtarış maşınlarını müxtəlif informasiya massivlərində və kolleksiyalarda paralel yerinə yetirməklə axtarış prosesinin tezləşdirilməsi;
- *avtonomluq* – sistemin uzaqda olan komponentlərini cəlb etmədən müstəqil şəkildə lokal axtarışların aparılması;
- *modulluluq* – axtarış maşınlarının işlənilib hazırlanmasının və fəaliyyətinin bir-birindən asılı olmaması, sistemin genişləndirilməsinin asanlıığı və s.;
- *sıradan çıxmaya davamlılıq* – bir axtarış maşınının və ya onun qoşulduğu rabitə kanalının sıradan çıxması digər axtarış maşınlarının, həmçinin bütövlükdə sistemin işinin dayanmasına səbəb olmur;
- *şəbəkə və zaman resurslarının qənaət olunması* və s.

Meta-axtarış sisteminin funksional struktur sxemi 8.2 sayılı şəkildə göstərilmişdir. Şəkildən göründüyü kimi, istifadəçi özünün informasiya tələbatına uyğun olaraq axtarış sorğusunu formalaşdırır və meta-axtarış sistemə ötürür.

Meta-axtarış sisteminin agenti aldığı axtarış sorğusunu müəyyən edilmiş axtarış maşınlarının formatlarına uyğunlaşdırır, müvafiq qaydada axtarış sorğularını tərtib edir və həmin axtarış serverlərinə ötürür. Həmin axtarış serverləri sorğuya uyğun axtarış aparır, nəticələri meta-axtarış sistemə göndərilir. Həmin nəticələr meta-axtarış sisteminin nəticələrin təhlili bloku tərəfindən ümumiləşdirilərək təhlil olunur və meta-axtarış sisteminin ümumi nəticəsi kimi istifadəçiyə təqdim edilir.



Şək.8.2. Meta-axtarış sisteminin funksional strukturu

8.4. İnternetin informasiya ehtiyatlarının indeksləşdirilməsi və təqdim olunması mexanizmləri

Qeyd olunduğu kimi, axtarış sistemlərində hər bir sənəd özünün axtarış obrazı ilə təqdim olunur. *Sənədin axtarış obrazı* (SAO) – informasiya ehtiyatlarının (sənədlərin) məzmununu tam əks etdirən təsviredicilər (terminlər, açar sözlər və s.) çoxluğunu özündə birləşdirir. SAO axtarış sistemində istifadə olunan model çərçivəsində informasiya massivinin ehtiyatlarının məzmunlarının indeksləşdirilməsi, yəni onların müəyyən olunmuş formaya inikas etdirilməsi prosedurasının tətbiqi nəticəsində yaranır.

Aydındır ki, SAO-nun yaradılması üçün sənədin məzmununa uyğun olan terminlər çoxluğunun müəyyənləşdirilməsi, sistemləşdirilərək saxlanması problemi həll olunmalıdır. Məlumdur ki, SAO-nun yaradılması, yəni sənədlərdən vacib terminlərin seçilməsi və onlara bağlanması prosedurası *sənədlərin indeksləşdirilməsi*, terminlər çoxluğu isə sənədin *indekslər bazası* və ya sadəcə *indeksi* adlanır.

Mümkün terminlərin sənədlərə bağlanmasını təmin etmək üçün təsbit olunmuş terminlər (informasiya massivinin sənədlərində rast gəlinə biləcək bütün terminlər) çoxluğunun – lüğətin yaradılması indeksləşdirmə ilə bağlı olan ən başlıca problemlərdən biridir. İndeksləşdirmə zaman sənədlərin məzmununu təsvir etmək üçün onun mətnindən götürülən və ya məzmununa uyğun olaraq rubrikatordan seçilən terminlər bu lüğətlə uzlaşdırılır.

İnternetdə hər gün yeni informasiya ehtiyatlarının yaranmasının (meydana gəlməsinin) və mövcud informasiyanın məhv edilməsinin mümkünlüyü səbəbindən İAS-da

indekslərin müntəzəm olaraq yeniləşdirilməsi, yəni sənədlərin təkrar indeksləşdirilməsi həyata keçirilməlidir. Bu məqsədlə, yuxarıda qeyd edildiyi kimi, axtarış robotlarından (robot proqramlarından) istifadə olunur.

Robot proqramı, bir qayda olaraq, “hörümçək”, “qurd”, “indeksləyici” adlanan və qarşılıqlı əlaqədə işləyən komponentlərdən ibarət olur.

“**Hörümçək**” (*spider*) əvvəlcədən verilmiş ünvanlar siyahısında göstərilən web-səhifələri axtarış maşınının müvəqqəti bazasına yükləyən proqramdır. O, istifadəçi kompüterinə yüklənmiş brauzer proqramına analoji qaydada işləyir, lakin web-səhifənin HTML-kodunu axtarış sistemə ötürür və monitorda heç məlumat əks etdirmir.

“**Qurd**” (*crawler*) “hörümçək” tərəfindən axtarış sistemə yüklənmiş web-səhifələrdə digər yeni web-səhifələrə istinadları (yəni yeni web-ünvanları) aşkar edərək axtarış sisteminin ünvanlar siyahısına əlavə edən proqramdır. Həmin ünvanlar siyahısı “hörümçəyin” hansı istiqamətdə hərəkət edəcəyini müəyyənləşdirir.

“**İndeksleyici**” (*Indexer*) “hörümçək” tərəfindən axtarış sistemə yüklənmiş web-səhifələri tərkib hissələrinə (başlıq, istinadlar, sənədin mətni) bölərək, onları, eləcə də sənəddə xüsusi stillərdə yazılmış (qalın, maili, altından xətt çəkilmiş şriftlərlə yazılmış, fonu rənglənmiş və s.) mətnləri təhlil edir. Sənəd, sənədin mövzusu üçün vacib olan, mətnə xüsusi çəki əmsalına malik elementlər – indekslər seçilir və sənədin axtarış obrazına daxil edilir.

Axtarış sistemlərində informasiya ehtiyatlarının indeksləşdirilməsi zamanı, əsasən, aşağıdakı mexanizmlərdən istifadə olunur:

I üsul. Sənədin təsvir edilməsi üçün məna təhlili əsasında onun kontekstini, yəni mahiyyətini və məzmununu daha dəqiq əks etdirən terminlər (açar sözlər) sistemdə mövcud olan xüsusi lüğətlərdən (məsələn, tematik rubrikator-dan) seçilir və SAO-ya daxil edilir. Bu zaman seçilən terminlər sənədin məzmununda rast gəlinməyə də bilər.

II üsul. Sənədin təsvir edilməsi üçün xüsusi üsulların köməyi ilə onun adından, başlığından və mətnindən, habelə onlarla bağlı digər xidməti məlumatlardan əhəmiyyət kəsb edən terminlər və açar sözlər seçilərək SAO-ya daxil edilir. Bu halda axtarış sistemində sinonimlər, assosiativ sözlər, stop-sözlər və digər lüğətlər mövcud olur ki, onların vasitəsilə SAO-ya yeni sözlər bağlanmış olur və genişləndirilir.

III üsul. Mövcud tematik rubrikatora uyğun olaraq sənədlərin məzmununun (kontekstinin) təhlili yolu ilə onların mətnindən həmin rubrikatorda olan müvafiq terminlər seçilir və SAO-ya daxil edilir.

Qeyd olunmalıdır ki, birinci yanaşma güclü semantik aparatın tətbiqini və ya indeksləşdirilən informasiya ehtiyatlarının tematikası sahəsində ixtisaslaşmış mütəxəssislərin iştirakı ilə əllə indeksləşdirmənin həyata keçirilməsini tələb edir. Əllə indeksləşdirmənin nəticəsi yüksək dəqiqliyi ilə seçilir, lakin yaxşı tematik kataloqun yaradılması personalın peşəkarlığından asılı olur. Bunun üçün böyük intellektual əmək, müvafiq mövzu sahələri üzrə dərin bilik və təcrübəyə malik olan mütəxəssislərin və ekspertlərin köməyi tələb olunur.

Semantik təhlilə əsaslanmış üsullar, əsasən, çətin reallaşdırılır və az inkişaf etmişdir. Lakin son dövrlərdə bu üsulların tətbiqi ilə tematik kataloqların avtomatik yolla, proqram təminatının köməyi ilə yaradılması məsələsi tədqiqat obyekt-

tinə çevrilmişdir. Bu istiqamətdə müəllif tərəfindən elmi araşdırmalar aparılmış, qeyri-səlis çoxluqlar və münasibətlər nəzəriyyəsinə əsaslanan üsullar təklif edilmişdir.

İnformasiya ehtiyatlarının məzmunundan daha əhəmiyyətli terminlərin seçilib çıxarılmasına və onların vaciblik dərəcələrinin hesablanmasına imkan verən üsullar *avtomatik indeksləşdirmə üsulları* adlanırlar.

Bu növ indeksləşdirmə üsullarını reallaşdıran axtarış sistemləri mütəxəssislərin və ekspertlərin iştirakı və müdaxiləsi olmadan işləyirlər. Yuxarıda qeyd olunduğu kimi, bu sistemlərin əsas komponentlərindən biri olan axtarış robotları internet şəbəkəsinin saytlarını "gəzir", bu saytlarda olan informasiya ehtiyatları haqqında məlumatları toplayır və onları xüsusi formada öz bazalarına daxil edirlər. Bu halda informasiya ehtiyatlarının tematik bölgüsü aparılmır və informasiya axtarışı yalnız açar sözlərə görə həyata keçirilir.

İndeksləşdirmə zamanı saytlarda olan web-səhifələrin hansının indeksləşdiriləcəyini axtarış robotları müəyyən edirlər. Lakin web-səhifələrin sahiblərinin arzusunu nəzərə almaq məqsədlə xüsusi standart – robotlar üçün istisnaetmə standartı istifadə olunur. Robotlar üçün istisnaetmə standartının mahiyyəti aşağıdakıdan ibarətdir: web-saytların administratorları öz serverlərində "robot.txt" faylı yaradır və bu faylda saytın web-səhifələrinin indeksləşdirilməsinə icazənin verilməsini və ya müxtəlif məhdudiyətlərin qoyulmasını göstərir. Başqa sözlə, bu faylda "kənar" robotlar tərəfindən saytın indeksləşdirilməsi siyasəti təsvir olunur.

Avtomatik indeksləşdirmə zamanı əlavə çətinlik yaranan aspektlər, yəni zəif cəhətlər aşağıdakılardan ibarətdir:

- indeksləşdirilən sənədlərdə müəlliflər tərəfindən mənası eyni (identik) olan anlayışların ifadəsi (təsviri) üçün müxtəlif terminlərdən istifadə oluna bilər. Məsələn, ingilis dilli mənbələrdə "informasiya axtarışı" termini bəzən "information retrieval", bəzən isə "information search" formalarında istifadə edilir;
- indeksləşdirilən sənədlərdə müvafiq sahədə geniş istifadə olunmayan və ya tematik rubrikatora daxil olmayan termin əhəmiyyətli açar söz kimi seçilə bilər, lakin bu terminin sinonimləri olan digər terminlər həmin sənəd üçün daha əhəmiyyətli ola bilər;
- bəzi terminlər tematika üzrə kifayət qədər yaxın olan başqa terminlərə assosiasiya yarada bilər;
- indeksləşdirilən sənədlərdə terminlər əvəzinə onların ingilis, latın və ya digər dillərdə ekvivalentləri istifadə oluna bilər;
- eyni tematika üzrə ayrı-ayrı sənədlərin dilləri müxtəlif ola bilər.

Yuxarıda sadalanan amilləri nəzərə almadan axtarışın nəticəsinin bilavasitə asılı olduğu indeksləşdirmə prosesinin arzu olunan səviyyədə keyfiyyətlə aparılmasını təmin etmək mümkün deyil. Burada sinonimlərdən və assosiativ sözlərdən, habelə onların başqa dillərə tərcümələrindən ibarət lüğətlərin yaradılması və istifadəsi vəziyyətdən çıxış yolu ola bilər.

Bundan əlavə, axtarış sistemlərinin istehsalçıları tərəfindən indeksləşdirmənin keyfiyyətini yaxşılaşdırmaq və əhəmiyyətsiz sözlərin indeks qismində qəbul olunmasının qarşısını almaq məqsədilə qadağan olunmuş ("stop-words"), ümumi təyinatlı, köməkçi və xidməti sözlərdən (bağlayıcılar, əvəzlilər, fellər, bəzi dillərdə sözünü şəkilçilər və s.) ibarət

lüğətlər istifadə olunur. Eyni zamanda indeksləşdirmə prosesində leksik vahidlərin normallaşdırılması həyata keçirilir.

Lüğətlərin və indekslər bazalarının süni surətdə şifirdilməməsi üçün sənədlərin axtarış obrazlarına daxil edilən terminlərə çəki əmsalları mənimsədilir. Bu əmsallar hər bir terminin verilən sənəd üçün əhəmiyyətlik dərəcəsini müəyyən edir.

Çəki əmsalları $[0,1]$ intervalında qiymət alır. Əgər əmsal 0 qiyməti alırsa, bu o deməkdir ki, termin sənəddə rast gəlinməmişdir, yəni termin sənəd üçün heç bir yaxınlıq (relevantlıq) münasibətinə malik deyildir. Əmsalın qiymətinin 1 olması isə terminin sənədin məzmununa tam (100%) uyğun gəldiyini göstərir.

Beləliklə, daha əhəmiyyətli terminlər, məsələn, 0.5-dən artıq çəkiyə malik olan terminlər, yəni çəki əmsalı $[0.5;1]$ intervalında olan terminlər sistemin indekslər bazasına daxil edilir.

Hazırda internet şəbəkəsində Google, Baidu, Bing, Yahoo, Excite, AltaVista, Lycos, Stars, Infoseek, Yandex, Rambler, Aport və s. kimi çoxlu sayda məşhur informasiya-axtarış sistemləri fəaliyyət göstərir. Bu sistemlərə qoşulan istifadəçilərə bilavasitə və ya dolayısı ilə internetin müxtəlif saytlarına giriş imkanı verilir. Onların köməyi ilə istifadəçilər öz sorğularını formalaşdırır, axtarışı həyata keçirir və tapılan sənədlərin, eləcə də web-səhifələrin URL ünvanlarını əldə edirlər.

Qeyd olunan axtarış sistemlərinin əksəriyyəti istifadəçilərə idman, kino, musiqi sahəsində məlumatlardan, kompüter oyunlarından, eləcə də elmi məqalələrdən, kitablardan, layihələrdən və s. kimi ehtiyatlardan ibarət olan müxtəlif sənədləri, web-səhifələri, faylları özündə saxlayan te-

matik kataloqlar təqdim edir. Avtomatik indeksləri reallaşdıran digər sistemlər isə açar sözlərə (terminlərə) əsasən dünyanın ayrı-ayrı ölkələrində yerləşən saytlardan müxtəlif məzmunlu sənədlərin axtarışını həyata keçirir. Bu sistemlər təqdim olunan xidmətlərin növlərinə, əhatə etdikləri mövzulara və saytlara, habelə indeksləşdirmə və kataloqlaşdırma alqoritmlərinə görə bir-birindən fərqlənirlər.

8.5. Axtarış spamları və onlara qarşı mübarizə

Axtarış spamı (spamdeksinq və ya web-spam) – axtarışın nəticəsində yüksək mövqe qazanmaq məqsədilə informasiya-axtarış sistemlərini aldatmaq və onun nəticələrini manipulyasiya etmək üçün çəhd edən web-saytlar və ya səhifələrdir. Axtarış spamları son nəticədə istifadəçilərin aldadılmasına yönəlmiş olur.

Axtarış spamı informasiya-axtarış sistemləri tərəfindən axtarışın nəticələrinə daxil edilməsini (yeridilməsini) təmin etmək məqsədilə məqsədyönlü şəkildə saytların qeyri-obyektiv (“qara”) optimallaşdırılması üçün nəzərdə tutulmuş geniş istifadə olunan alətdir.

Bu gün saytların internetdə irəliləməsi (daha geniş tanınması) məqsədilə axtarış spamlarının tətbiqinə qadağa qoyulmasına baxmayaraq, praktiki olaraq, bu texnologiyadan çox istifadə olunur. Axtarış spamlarını düzgün istifadə edilməsi qısa zamanda web-saytın tanınmasına və ona müraciətlərin həcmnin artmasına kifayət qədər əhəmiyyətli təsir göstərə bilər.

Bir çox hallarda yeni fəaliyyətə başlayan web-sayt mütəxəssisləri bilmədən axtarış spamının bu və ya digər

üsullərindən istifadə edirlər. Axtarış spamının bir növü qismində URL istinadların siyahısının alınmasını (satin alınmasını) göstərmək olar. Qeyd etmək lazımdır ki, bir çox hallarda belə axtarış spamları saytlara axtarış sistemləri tərəfindən süzgeclərin (məhdudiyyətlərin) qoyulmasına gətirib çıxarır.

Saytların optimallaşdırılması üçün istifadə olunan kifayət qədər çoxlu sayda axtarış spamları mövcuddur. Axtarış spamlarını əsas iki qrupa bölmək olar:

- kontekst (məzmun) və ya mətn spamları;
- istinad spamları.

Kontekst spamları dedikdə mahiyyət etibarlı ilə saytın və ya web-səhifənin əsas mətninə açar sözlərin və ifadələrin əlavə edilməsi, həddən artıq doldurulması, məzmunun qarışdırılması başa düşülür. Lakin kontekst spamlarının digər formalarına da rast gəlinir: məsələn, web-səhifəyə fonla eyni rəngdə, şriftinin ölçüsü 1 piksel olan (gözlə görünməyən) mətnin, səhifəyə açar sözlər olan teqlərin daxil edilməsi.

Kontekst spamların yaradılması üçün bir neçə üsuldən istifadə olunur:

- *Açar sözlərin təkrarlanması.* Axtarış spamı saytın məzmununa daxil olan müxtəlif informasiya elementlərində (səhifənin mətnində, başlıqda, meta-teqlərdə, istinadlarda və s.) açar sözlərin çoxsaylı təkrarlanması yolu ilə yaradılır.
- *Müxtəlif açar sözlərin yığının yaradılması.* Web-səhifə çoxlu sayda müxtəlif açar sözlər çoxluğu (yığını) ilə doldurulduqda, müəyyən az tezlikli sorğuların ona daxil olması ehtimalını artırmaqla axtarış spamını formalaşdırılır.

- *Açar sözlərin səhifənin məzmununa salınması.* Digər informasiya mənbələrindən sitatlar gətirməklə lazımi açar sözlər və ya ifadələr səhifənin məzmununa salınır.
- *Məzmunların qarışdırılması.* Unikal məzmunun yaradılması məqsədilə məzmunu müxtəlif mənbələrdən götürülmüş səhifələr qarışdırılıraq sayta daxil edilir.

Qeyd etmək lazımdır ki, bu gün bir çox informasiya-axtarış sistemləri kontekst spamlarını kifayət qədər effektiv şəkildə aşkar edə və onlara qarşı mübarizə apara bilər. Belə ki, kontekst spam istifadə edən saytlara qarşı süz-gəclər qoyulur və ya axtarış alqoritmləri tərəfindən əks mənada nəzərə alınır.

İstinad spamları axtarış spamlarının daha geniş yayılmış növüdür. İstinad spamlarının mahiyyəti müxtəlif internet ehtiyatlarına istinadların istifadəsində süni manipulyasiya edilməsindən ibarətdir. Burada spam rolunda həm daxili, həm də xarici istinadlar çıxış edə bilər. Xarici istinadlar qismində istinadlar kataloqundan, ələxüsus qeyri-mövzu kataloqlarından istifadə olunur.

Daxili istinadlara isə verilmiş sayta istinadların yerləşdirilməsi üçün xüsusi saytların yaradılması, donor-saytda gizli istinadın qoyulması, istinadların mübadiləsi sistemlərindən istifadə, daxili informasiya ehtiyatlarının mübadiləsinin təşkili, kommentarılərdə istinadların qoyulması, boşaldılmış domenlərin alınması yolu ilə onlara daxil olan və onlardan çıxan istinadların istifadəsi, nizamlanmayan istinadların (qonaq kitablarında, elanlar lövhələrində və s.) qoyulması və digər mexanizmlər aid edilir.

İstinad spamları ən təhlükəsiz axtarış spamı hesab olunur. Belə ki, axtarış sistemləri saytın bu və ya digər istinadının məhz necə yarandığını heç də həmişə dəqiq müəyyən etmək imkanına malik olmur. Lakin istinad spamının həddən artıq istifadəsi ciddi fəsadlarla (süzgəcin və ya məhdudiyətin qoyulması) nəticələnə bilər. Ona görə də saytın istinad çəkisini alınmış istinadlar vasitəsilə dolduran zaman nəzərə almaq lazımdır ki, bu istinadlar həmin sayt üçün təbii əlamətləri özündə ehtiva etməli, mövzu sahəsinə aid olmalıdır və s.

Praktikada kontenks və istinad spamları ilə yanaşı kloaking, redirekt, gizli məzmun, dorvey (doorway – giriş səhifələri) və svopinq (swapping) kimi spamlardan da istifadə olunur.

Kloaking – sayta baxan istifadəçiyə bir web-səhifəni, axtarış sisteminin robotuna isə başqa səhifəni təqdim edən “qara” axtarış optimallaşdırması (spamı) üsuludur.

Redirekt dedikdə istifadəçinin bir saytdan digərinə avtomatik olaraq (refresh meta-teqi, cgi-ssenari, Javascript və ya server əlavəsi vasitəsilə) istiqamətləndirilməsi başa düşülür. Başqa sözlə, istifadəçi brauzerdə hər hansı bir ünvan yığır, lakin ekranda tamamilə başqa ünvana malik sayt açılır.

Gizli məzmun (kontent) – web-səhifədə mövcud olan, lakin istifadəçiyə görünməyən mətdir. Səhifədə mətnin gizlədilməsi, mətnin ekranda sürüşdürülməsi (fırlanması), səhifələrin təkrarlanması, CSS (Cascading Style Sheets – stillərin kaskad cədvəlləri), Javascript, bir layın digər layın üzərinə qoyulması və s. üsullar vasitəsilə həyata keçirilir.

Dorvey – konkret axtarış sistemində konkret axtarış soğusuna uyğun xüsusi şəkildə optimallaşdırılmış web-

səhifədir. Hər bir açar sözə və ya ifadə, eləcə də hər bir axtarış maşını üçün ayrı-ayrı səhifə (və ya səhifələr) işlənilib hazırlana bilər. Bir qayda olaraq, web-səhifələr 1-3 açar sözə və ya ifadəyə görə optimallaşdırılır. Dorvey əsas saytın hissəsi hesab olunmur. Onun məqsədi səhifəyə daxil olanları əsas sayta istiqamətləndirməkdən ibarətdir.

Svopinq – axtarış sistemi tərəfindən təqdim olunan axtarışın nəticəsində yuxarı yerlərin əldə olunması üçün saytın web-səhifələrinin həmin axtarış sistemi və konkret axtarış sorğusu üçün xüsusi şəkildə optimallaşdırılmasıdır. Sonradan web-səhifənin uğurlu indeksləşdirilməsi və lazımi reytingin qazanılması əldə olunduqdan sonra həmin səhifənin məzmunu tamamilə və ya qismən dəyişdirilir və ya başqa web-səhifə ilə əvəz olunur.

Məlumdur ki, axtarış spamlarının tətbiqi axtarış sistemlərinin səmərəliliyinə və axtarışın nəticəsinə, yekunda isə axtarış aparən insanlara mənfi təsir göstərir. Axtarış spamlarının yaratdığı əsas problem lüzumsuz (“zibillənmiş”) kontentin yaradılması ilə bağlıdır ki, bu da axtarış sistemlərinin səmərəli işini çətinləşdirir, axtarışın nəticəsində verilən internet ehtiyatların obyektiv nizamlanmasının və relevantlıq dərəcələrinin təhrif olunmasına gətirib çıxarır.

İlk dövrlərdə axtarış sistemləri web-səhifələrdə və onların kodlarında rast gəlinən açar sözlərə etibar edərək tezlikləri hesablayırdılarsa, hazırda əksər axtarış maşınları səhifələri ciddi təhlil edir və süni şəkildə daxil edilmiş açar sözləri aşkarlamağa çalışır.

Axtarış spamlarına qarşı mübarizə aparmaq məqsədi ilə axtarış sistemləri tərəfindən müxtəlif üsullar işlənilib hazırlanır və antispam proqramlarından istifadə edilir.

IX FƏSİL

İNTERNETİN MƏŞHUR İNFORMASIYA-AXTARIŞ SİSTEMLƏRİ

**İnternetin informasiya-axtarış sistemlərinin yaranma
və inkişaf tarixi**

Google axtarış sistemi haqqında ümumi məlumat

Google axtarış sisteminin inkişaf tarixi

Google sisteminin axtarış alqoritmləri

Google axtarış sisteminin əsas məhsulları

Google sisteminin axtarış imkanları

Google axtarış sisteminin interfeysi

**Google-da axtarış sorğusunun tərtib edilməsi
prinsipləri və şərtləri**

9.1. İnternetin informasiya-axtarış sistemlərinin yaranma və inkişaf tarixi

İnternetdə dünya hörümçək toru (WWW) yaranana qədər ilk axtarış alətlərindən biri *Archie* olmuşdur. Archie Kanadanın McGill Universitetinin əməkdaşları Alan Emteyc, Piter Doyç və Bil Heelan tərəfindən 1990-cı ildə işlənib hazırlanmışdır. Bu proqram FTP-ehtiyatlarının siyahısını emal edir və faylların adlarına görə axtarış aparmağa imkan verir.

Web-fəza üçün ilk axtarış sistemi *Wandex* Massaçuset Texnoloji İnstitutunun əməkdaşı Metyu Qrey tərəfindən 1993-cü ildə yaradılmışdır. Onun əsasını təşkil edən “World Wide Web Wanderer” proqramı internet səhifələri indeksləşdirən ilk robot olmuşdur. Wandex məhz bu robot tərəfindən yaradılan indexlər bazasının adı idi. Lakin zaman keçdikcə bu indexlər bazası istifadəçilər üçün aktuallığını itirdi və axtarış üçün daha az yararlı oldu.

1993-cü ildə Martin Koster tərəfindən yaradılan və web-fəzanın ilk tematik kataloqu olan *Aliweb* axtarış sistemi fəaliyyətə başladı. Bu kataloqun bazasını pobot proqramı deyil, saytların və ya səhifələrin sahibləri doldururdular. Aliweb axtarış sistemi internetdə indiyədək fəaliyyət göstərir.

1993-cü ildə Nevada Universiteti tərəfindən Gopher serverlərinin sənədləri (mətnləri) üzrə axtarış aparən *Veronica* sistemi istifadəyə verilmişdi. Həmin il *Jump Station*, *WWW Worm*, *RSBE* (Repository-Based Software Engineering) axtarış sistemləri də fəaliyyət göstərməyə başlamışdılar.

İnternetdə informasiya axtarışını həyata keçirən *Web-Crawler*, *Excite*, *Infoseek*, *Yahoo*, *Tradeware Galaxy* və s. axtarış sistemləri 1994-cü ildən fəaliyyətə başladılar.

İnternetin ilk tammətnli axtarış sistemi olan **WebCrawler** 1994-cü ildə America Online şirkəti tərəfindən yaradılmışdı. Onun robot proqramı internetin informasiya ehtiyatlarını indeksləşdirirdi. Əvvəlki axtarış sistemlərindən fərqli olaraq, WebCrawler istifadəçilərə istənilən web-səhifədə istənilən açar sözə görə axtarış aparmağa imkan verirdi. Axtarış prosesi üçün bu sistemdə reallaşdırılan belə yanaşma sonradan internetin əsas axtarış sistemləri üçün standart kimi qəbul edildi. Qeyd edilməlidir ki, WebCrawler axtarış sistemi geniş dairələrdə məşhur olan ilk axtarış sistemi idi. Sonradan WebCrawler sistemi Excite şirkəti tərəfindən alınmışdır.

IAC Search & Media şirkətinə məxsus olan **Excite** axtarış sisteminin (URL: <http://www.excite.com>) əsası 1994-cü ildə Stendfor Universitetinin tələbələri Qrem Spenser, Co Kraus, Mark Van Xaren, Rayan Makintayr, Ben Luç və Martin Reynfrid tərəfindən qoyulmuşdur. Excite sisteminin axtarış mexanizmləri standart axtarış mexanizmləri (açar sözlərə görə axtarış) ilə yanaşı, məzmunu görə evristik axtarış üsullarından, eləcə də sorğuların tərtib edilməsi zamanı məntiqi əməliyyatlardan istifadə edirdi.

Infoseek axtarış sisteminin əsası Stiv Kirs tərəfindən 1994-cü ildə qoyulmuşdur. İlk dövrlərdə tam bazadan istifadə kommersiya əsasında həyata keçirilirdi, lakin sonradan bu xidmət pulsuz oldu. Hazırda Infoseek ən universal axtarış sistemlərindən biridir. Belə ki, o, web-səhifələr, Usenet məlumatlar, mövzular üzrə xəbərlər, şirkətlər və korporativ xəbərlər, xəritələr və s. üzrə axtarış, eləcə də adamların axtarışı kimi xidmətləri təklif edir. Eyni zamanda Infoseek öz istifadəçilərinə keyfiyyətli mövzu kataloqunu təqdim edir. İstifadəçilər üçün əlverişli interfeys

reallaşdırılmışdır. Infoseek axtarış maşınının indekslər bazasında “ölü” (işləməyən) istinadların silinməsi mexanizmi fəaliyyət göstərir.

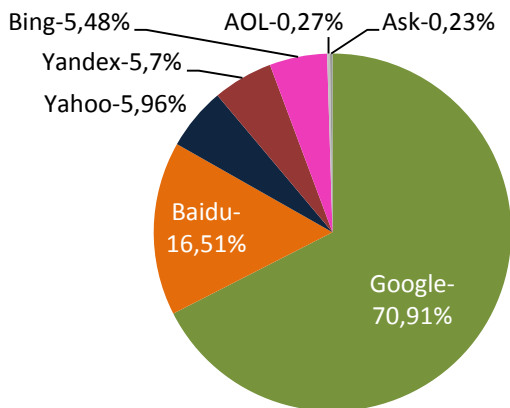
Yahoo axtarış sistemi 1994-cü ilin əvvəlində Stendford Universitetinin aspirantları Devid Faylo və Cerri Yanq tərəfindən dünya hörümçək toru üzrə bələdçi sayt kimi yaradılmışdır. Bu sayt digər saytların kataloqunu özündə əks etdirirdi. Yahoo dünyanın ən məşhur axtarış sistemlərindən biridir. Bu gün Yahoo özündə Yahoo!Directory tematik kataloqunu, eləcə də internetin ilk və ən məşhur elektron poçtu olan Yahoo!Mail xidmətlərini birləşdirir.

1995-ci ildə Lycos (Karneqi Melon Universiteti), AltaVista (AltaVista Technology şirkəti, Lui Mone və Maykl Berrouz – sonradan DEC – Digital Equipment Corporation şirkətinə satılmışdır), MetaCrawler (Eric Selburg – Washington University) sistemləri internet istifadəçilərinin istifadəsinə verildi.

Lycos axtarış sistemi Karneqi Melon Universitetində Maykl Loren Molden tərəfindən aparılan tədqiqatların nəticəsi olaraq 1995-ci ildə meydana gəldi. Həmin il Lycos şirkəti də yaradıldı və fəaliyyətə başladı. Lycos internetin ilk axtarış maşınlarından biridir. Bu gün onun axtarış mexanizmləri böyük sayda axtarış alətlərindən istifadə etməklə açar sözlərə və bütöv ifadələrə görə tammətnli axtarışı həyata keçirir. O, web-səhifələrin URL-ünvanlarına istinadları, telekonfranslarda, FTP-serverlərdə və arxivlərdə, eləcə də internetin digər xidmətləri vasitəsilə əldə oluna bilən informasiya ehtiyatları, o cümlədən elektron poçtu ünvanları üzrə axtarış aparmağa imkan verir. Lycos sisteminin 15-dən çox avropa dillərində interfeysi təqdim edir.

1990-cı illərin ikinci yarısında internetin məşhur axtarış sistemləri olan Inktomi (Kaliforniya Universiteti – Erik Bryuyer və Pol Qotye, 1997-ci il), AskJeeves (Qarri Qrune və Devid Ueven, 1997-ci il), Overture (Overture, Idealab – Bill Qross, 1998-ci il), Google (Stenford Universiteti, Larri Peyc və Sergey Brin, 1999-cu il) öz xidmətlərini təqdim etməyə başladılar.

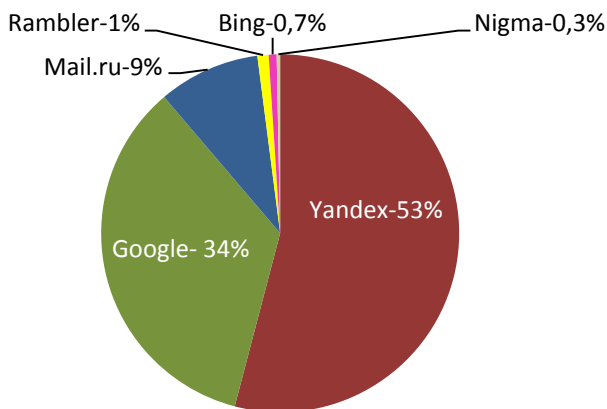
2014-cü ilin əvvəlinə olan statistik məlumatlara görə bu gün dünyanın ən tanınmış və geniş istifadə olunan axtarış sistemi Google sistemidir. O, dünyada informasiya axtarışı bazarının 70,91%-ni təmin edir. İkinci yeri Çin axtarış sistemi olan Baidu (18,51%) tutur. Növbəti yerlərdə Yahoo, Yandex, Bing, AOL və Ask axtarış sistemləri gəlir (şək.9.1).



Şək.9.1. Dünyanın ən məşhur axtarış sistemləri

Bu göstəricilər çox dinamikdir, coğrafi və demiqrafik məkanlardan asılı olaraq fərqlənə və zaman keçdikcə dəyişə bilər.

İnternetin rus dilli məkanında axtarış sistemlərinin reytingi 9.2 sayılı şəkildə verilmişdir.



Şək.9.2. Rus dilli məkanda ən məşhur axtarış sistemləri

Azərbaycanda da ən çox istifadə olunan axtarış sistemi Google sistemidir. İstifadə edənlərin sayına görə ikinci yeri Yandex axtarış sistemi tutur.

9.2. Google axtarış sistemi haqqında ümumi məlumat

Google – dünyada internetin ən məşhur, ən çox istifadə olunan və ən geniş yayılmış informasiya-axtarış sistemidir.

“Google” sözü 10^{100} , yəni arxasında 100 sıfır olan 1 ədədini ifadə edən riyazi termin olan “Googol” sözündən götürülmüşdür. “Googol” termini amerika riyaziyyatçısı Edvard Kasnerin bacısı oğlu Milton Sirotto tərəfindən təklif edilmişdir. Onun bu təklifi ilk dəfə Edvard Kasner və Ceyms Nyumenin “Riyaziyyat və təsəvvür” (“Mathematics and the Imagination”) adlı kitabında verilmişdir.

Bu terminin istifadəsi Google şirkətinin və onun axtarış sisteminin internetin nəhəng, praktiki olaraq, sonsuz infor-

masiya massivində əlverişli axtarışın təşkilinə yönəlmiş məqsədini özündə əks etdirir.

Bu gün Google axtarış sistemi eyni adlı şirkət tərəfindən idarə olunur. Google şirkəti öz imkanlarını ideal axtarış serverinin işlənilib hazırlanması istiqamətində cəmləşdirən şirkətdir. Şirkətin təsisçilərinin fikrinə görə Google axtarış serveri istifadəçilərə nə lazım olduğunu başa düşür və onlara məhz istədiklərini təqdim edir. Google şirkətinin missiyası qlobal informasiya fəzasının (web-fəzanın) əlverişli təşkilindən, onun hərtərəfli əldə edilən və faydalı olmasından ibarətdir. Bu məqsədlə Google şirkəti daim yeni texnologiyalar işləyib hazırlayır və tətbiq edir.

Google sisteminin saytında axtarış modulu ilə yanaşı adi internet istifadəçiləri və web-masterlər üçün əlverişli olan kifayət qədər müxtəlif informasiya-axtarış xidmətləri, əlavə imkanlar, o cümlədən saytların daha geniş tanıtılması və kommersiya məqsədilə kontekstə uyğun reklamın yerləşdirilməsi və web-nəşrlərin təşkili üçün digər funksiyalar reallaşdırılmışdır.

Google reklamçılar üçün Google AdWords və web-menecerlər üçün Google AdSense proqramlarını təqdim edir. AdWords – reklamçılara məqsədyönlü reklam vasitəsilə öz məhsullarını və internet xidmətlərini önə çəkmək üçün istifadə olunan proqramdır. AdSense – web-menecerlər tərəfindən öz saytlarında relevant məzmunlu reklamların göstərilməsi və qazanc əldə edilməsi imkanlarını yaradan proqramdır.

Bu texnologiyalara əsasən Google reklam şirkətlərinə internetdə effektiv və sərfəli reklam yaymaq imkanını təqdim edir. Bu zaman reklamın mövzusu istifadəçiyə təqdim edilən səhifədə əks olunan informasiyanın məzmununa relevantliq təşkil edir. Ona görə də, reklam yalnız onu

yerləşdirənlər üçün sərfəli deyil, həm də istifadəçilər üçün faydalı olur.

Google şirkətinin prinsiplərinə görə istifadəçi reklamda olan məlumatların axtarışın nəticəsinin təqdim edildiyi səhifədə əks olunması (yəni istifadəçinin bu reklamı görməsi) üçün kiminsə pul ödədiyini bilmək hüququna məlikdir. Ona görə də reklam (elan) axtarışın nəticəsindən və ya səhifənin əsas məzmunundan ayrılmalıdır. Bu baxımdan Google axtarış nəticəsində istifadəçiyə təqdim olunan informasiya içərisində reklam üçün yer satmır və reklamçılara pul ödəmək yolu ilə axtarışın nəticəsində daha yüksək sıralarda yer tutmağa imkan vermir.

Google sisteminin rahat və sadə olması onu dünyanın ən məşhur axtarış serverlərindən birinə çevirmişdir.

Google xüsusi alətlər paneli vasitəsilə saytın əsas səhifəsinə daxil olmadan internetin informasiya fəzasına çıxaraq axtarış aparmağa imkan verir. Google Deskbar modulu isə Google sisteminin axtarış pəncərəsini Windows sisteminin məsələlər panelinə yerləşdirir. Onun vasitəsilə brauzer proqramını açmadan istifadə olunan istənilən Windows əlavəsindən internetdə axtarış aparmaq mümkündür. Hətta WAP və i-mode rejimlərində işləyən müxtəlif mobil platformalardan, o cümlədən mobil telefonlardan istifadə etməklə Google axtarış serverinin imkanlarından faydalanmaq mümkündür.

Google axtarış sistemi müxtəlif dillərdə tərtib edilmiş sorğulara görə HTML-sənədlərlə yanaşı Adobe Acrobat PDF, Microsoft Office, PostScript, Corel WordPerfect, Lotus 1-2-3 və s. sənədlər, şəkil (qrafika), xəritələr, audio-video məlumatlar, xəbərlər (o cümlədən Usenet-xəbərlər), müxtəlif dillərdə olan mənbələr üzrə sürətli axtarış aparmağa imkan verir.

Google istifadəçilərinə mətnlərin 81 dilə tərcüməsini həyata keçirən tərcümə sistemini təklif edir. Google sisteminin təklif etdiyi dillərin sayı daim artır.

9.3. Google axtarış sisteminin inkişaf tarixi

1995-ci ildə Stenford Universitetinin kompüter elmləri kafedrasının iki aspirantı – Larri Peyc və Sergey Brin informasiya texnologiyaları sahəsinin ən ciddi problemlərindən biri olan böyük məlumatlar massivində (o cümlədən internetdə) lazımı informasiyanın axtarılması məsələsinin həllinə vahid yanaşmanı formalaşdırdılar. Bu yanaşma hər hansı sayta internetin digər informasiya ehtiyatlarından istiqamətlənmiş istinadların təhlilinə əsaslanan unikal yanaşma idi. Əsası 1996-cı ilin yanvarında qoyulan və BackRub adlanan bu texnologiya 1997-ci ildə artıq kifayət qədər tanınmışdı.

1998-ci ilin birinci yarısında ümumi həcmi 1 terabayt olan bərk disklər alan L.Peyc və S.Brinq Google sisteminin ilk verilənlər mərkəzini (bazasını) yaratdılar. Həmin il Google şirkəti yaradıldı və Kaliforniyada Menlo-Parkda fəaliyyətə başladı. Həmin dövrdə Google.com saytı (saytın beta versiyası) artıq gün ərzində 10000 axtarış sorğusunu emal edirdi. Həmin ilin dekabrında isə (o zaman şirkət qarajda yerləşirdi və 3 əməkdaşı vardı) Google saytı PC Magazin jurnalı tərəfindən dünyanın 100 ən yaxşı veb-saytları və axtarış serverləri sırasına daxil edilmişdi.

1999-cu ilin əvvəlində Palo-Alto şəhərində Avenyu Universitetində əsas ofisinə köçən, artıq 8 əməkdaşı olan və gün ərzində 500000 sorğu emal edən Google Inc. Şirkətinə maraq günbəgün artırdı.

ABŞ-ın aparıcı şirkətlərini investorlar qismində cəlb etməyə başlayan Google şirkəti 1999-cu ilin iyun ayında Si-likon vadisinin aparıcı Sequoia Capital və Kleiner Perkins Caufield & Byers şirkətlərindən 25 milyon ABŞ dolları həcmində kredit ala bildi. Həmin dövrdə Google Auntin-Vyuda (Kaliforniya) yerləşən Googleplex adlanan ofisə (bu ofis hazırda da şirkətin mənzil-qərargahıdır) köçdü.

AOL/Netscape şirkətinin, Virgilo İtaliya portalının və Virgin Net Böyük Britaniya soraqçasının Google-un müştəriləri sırasına qoşulması nəticəsində ona müraciət edilən gündəlik axtarış sorğularının sayı 3 milyonluğa çatdı.

1999-cu ilin sentyabr ayının 21-də Google web-saytından “beta” əlavəsini götürdü. Google 1999-cu ildə Time jurnalı tərəfindən dünyanın ən yaxşı kibernetik texnologiyalar onluğuna daxil edildi. İnternetin milyarddan artıq səhifəsini indeksləşdirməklə dünyanın ən böyük axtarış serveri kimi tanınan Google 2000-ci ildə texniki nailiyyətlərə görə Webby Award və People’s Voice Award mükafatlarına layiq görüldü. Həmin ilin sonunda Google gün ərzində 100 milyon axtarış sorğusu emal edirdi.

2001-ci ilin əvvəlində Deja.com xəbərlər arxivinə sahib olan Google internetin ən böyük xəbərlər sistemi Usenet arxivinin axtarış formatına çevrilməsi kimi çətin məsələni həyata keçirdi və uzun illər ərzində toplanmış 500 milyondan artıq xəbəri indeksləşdirdi. Həmin ilin sonunda Google axtarış sisteminin indeksi 3 milyard sənədi özündə birləşdirirdi.

Lycos Koreya şirkəti ilə imzalanmış razılaşma Asiyada daha geniş istifadəçi dairəsi üçün Google axtarış imkanları yarandı, Universo Online (UOL) ilə əməkdaşlıq işə Google axtarış sistemini Latın Amerikasında əsas axtarış

serverinə çevirdi. Google artıq 28 dildə, o cümlədən türk və ərəb dillərində axtarış aparmağa imkan verirdi.

2002-ci ilin fevralında Google Search Appliance aparat axtarış kompleksi tətbiq olunmağa başladı. Asanlıqla qoşula bilən bu aparat vasitəsi korporativ intraşəbəkələrin, elektron ticarət saytlarının, universitetlərin və müxtəlif səviyyəli təşkilatların şəbəkələrinin skan edilməsini həyata keçirirdi. Həmin dövrdə 2001-ci il üçün Search Engine Watch Awards konkursunda Web-masterlər cəmiyyəti Google şirkətinə bir sıra nominasiyalar üzrə prizləri təqdim etmişdi:

- qabaqcıl axtarış xidmətinə görə – Outstanding Search Service;
- təsvirlər (şəkillər) üzrə ən yaxşı axtarış serveri – Best Image Search Engine;
- ən yaxşı dizayn – Best Design;
- web-masterlər üçün ən rahat axtarış sistemi – Most Webmaster Friendly Search Engine;
- ən yaxşı axtarış funksiyası – Best Search Feature.

Google-un internet əlavələri (API) proqram interfeysinin qurulması digər proqramlara axtarış sisteminə bilavasitə sorğular göndərmək və milyardlarla web-sənədlərdən məlumatlar əldə etmək, habelə Google-un bazasında oyundan başlayaraq yeni axtarış interfeyslərinədək müxtəlif yenilikləri və texnologiyaları tətbiq etmək imkanlarını yaratdı.

2002-ci ildə Google AdWords reklam sistemini, o cümlədən reklam və elanlar üzərində hər “klikə” görə qiymətin hesablanması sxemini yeniləşdirdi. Bu sxemə görə axtarış səhifələrində reklam yerləşdirmək həm böyük, həm də kiçik müəssisələr üçün iqtisadi baxımdan sərfəli oldu.

Google-un reklam siyasətində əsas diqqət istifadəçinin üzərində cəmləşir. Reklam səhəsində yenilik olan bu

yanaşma istifadəçilərin sorğularının məzmununa uyğunluq, reklamın göstərilməsi məqsədilə müvafiq auditoriyanın seçilməsi üçün açar sözlərin istifadəsi və elanların reytinginin müəyyən edilməsi texnologiyalarını özündə birləşdirir. Bu yanaşmanın sayəsində elanlar yalnız onları maraqlandıra biləcək istifadəçilərə göstərilir. Ona görə də Google-nin reklam siyasəti həm istifadəçilər, həm də reklamverənlər üçün sərfəli olur.

AdWords reklam sistemi elə etimad qazandı ki, 2002-ci ilin mayında Amerikanın internet xidməti təqdim edən əsas şirkəti olan America Online özünün 34 milyonluq istifadəçi və on milyonlarla digər müştəriləri üçün axtarış və reklam sistemi qismində Google axtarış sistemini seçmişdi. Həmin dövrdə America Online Google sistemini web-axtarışın çempionu və kralı adlandırmışdı.

Sentyabr ayında Google Xəbərlər xidmətinin beta-sersiyası, dekabr ayında isə Froogle xidməti fəaliyyətə başladı. Xəbərlər xidməti eyni zamanda dünyanın 4500 aparıcı xəbər mənbələrinə girişi təmin edirdi. Froogle xidməti milyonlarla relevant web-saytlar arasından müəyyən məhsulların bir neçə mənbəsini tapmaqda (müvafiq şəkil və qiymətləri əks etdirməklə) istifadəçilərə yardım göstərir.

2003-cü ildə Google Pyra Labs şirkətini aldı və Blogger evinə çevrildi. Bir müddət sonra Google AdSense proqramını reallaşdırıldı. AdSense istənilən ölçüdə web-saytlara sadə üsulla (web-səhifələrdə onların məzmununa uyğun elanların yerləşdirilməsi yolu ilə) qazanc əldə etməyə imkan verirdi. Belə ki, Google AdSense texnologiyası istənilən web-səhifənin mətnini təhlil edir və ona uyğun olan elanları göstərir. Google saytının indeksinin həcmi 4,28 milyard web-səhifə olduğu həmin ildə Amerika marketing jurnalı

Brandchannel yenidən Google şirkətini ilin brendi, ABC News isə L.Peyci və S.Brini “Həftənin adamları” adlandırdı.

2004-cü ilin fevralında Google indeksi 6 milyarddan artıq elementin, o cümlədən 4,28 milyard web-səhifənin, 880 milyon şəklin, 845 milyon Usenet xəbər və kitablara dair səhifələrin kolleksiyasını özündə birləşdirirdi.

Həmin ilin aprel ayında Google şirkəti elmi-tədqiqat mərkəzinin və Gmail poçt xidmətinin açılması planı barədə məlumat yaydı. Gmail poçt xidməti hər bir istifadəçiyə həcmi 1 GB olan pulsuz poçt qutusu təqdim edirdi.

May ayında Google məşhur onlayn nəşr xidməti olan Blogger-in yeni versiyasını, oktyabrında isə pulsuz yüklənən və fərdi kompüterdə faylların lokal axtarılması üçün nəzərdə tutulan Google Desktop Search əlavəsini, eləcə də Google SMS xidmətini istifadəyə verdi. Sentyabr ayında artıq Google saytının müxtəlif ölkələrdə 100-dən artıq domeni (axtarış serveri) fəaliyyət göstərirdi. Oktyabr ayında Google rəqəmsal və peyk xəritələrinin işlənməsi ilə məşğul olan Keyhole Corp şirkətini aldı və istifadəçilərə Yer kürəsinin istənilən nöqtəsinin üçölçülü təsvirinə baxmaq üçün yeni güclü axtarış imkanı təqdim etdi. Noyabr ayında Google-un indeksinin həcmi 8 milyard web-səhifəni ötüb keçmişdi.

2005-ci ildə korporativ şəbəkələrdə daxili saytlar və şəbəkələr üzrə axtarış aparmaq üçün nəzərdə tutulmuş Google Mini, televiziya verilişlərindən informasiya yazmağa və onlar üzrə axtarış aparmağa imkan verən Google Video sistemləri istifadəyə verildi.

Həmin dövrdə Google-un təsvirlər (şəkillər) üzrə axtarış xidmətinin verilənlər bazasına 1 milyarddan artıq təsvir (fotoqrafiyalar, çertyojlar, şəkillər, eskizlər, karikaturalar, plakatlar və s.) daxil idi. Avqust ayında kompüter vasitəsilə istənilən vaxtda və istənilən yerdə insanlar arasında

ünsiyyət yaratmaq üçün nəzərdə tutulmuş pulsuz vasitə olan Google Talk, səsin ötürülməsi üçün Crystal-Clear texnologiyaları, sentyabr ayında bütün dünya üzrə bloqlarda sürətli informasiya axtarışını təmin edən Google Blog Search modulu istifadəyə verildi.

Noyabr ayında müxtəlif məzmunlu məlumatların (məsələn, malların siyahısı, elmi məlumatlar, reseptlər, siyahılar və s.) yüklənməsini təmin edən Google Base, onlayn və marketinq şirkətlərinin faktiki nəticələrini təhlil etmək imkanı verən Google Analytics xidmətləri reallaşdırıldı. İlin sonuna doğru Google-da özündə bəstəkarlar, ifaçılar, musiqinin adı, janrı, alboma istinadlar, rəylər, onların satıldığı mağazalar haqqında məlumatları birləşdirən musiqi faylları üzrə axtarış funksiyası işə başladı.

2005-ci ildə peykdən alınmış təsvirlər şəklində xəritələrə baxmaq, üç ölçülü qurğu və tikililəri göstərmək imkanını reallaşdıran, kartoqrafik və axtarış funksiyalarını təqdim edən Google Yer Planeti xidməti təqdim olundu. Windows mühitində mətn və səs tipli məlumatları mübadilə etməyə imkan verən Google Talk proqram əlavəsi istifadəçilərə təqdim olundu. Həmin il Google sisteminə maşın tərcüməsi funksiyası əlavə edildi.

2006-cı ildə Google PageRank texnologiyasını reallaşdırdı. PageRank web-səhifələrin əhəmiyyətliliyini obyektiv qiymətləndirməyə imkan verirdi. Bunun üçün PageRank modulu 500-dən çox dəyişənə və 2 milyarddan artıq terminə görə tənlikləri həll edir. PageRank texnologiyasına görə birbaşa istinadlar əvəzinə bir səhifənin digər səhifəyə istinadı ikinci səhifənin xeyrinə (və ya zərərinə) səsvermə kimi qəbul edilir və verilmiş səsələrin sayına görə konkret səhifənin əhəmiyyətliliyi qiymətləndirilir.

Google PageRank texnologiyası ilə yanaşı hipermətnlərin məzmununu təhlil edir, şriftlərin ölçüsü, stili və forması, altbölmələrin adları, sözlərin sənədlərdə yeri və s. nəzərə alınır.

Google simsiz qurğularda axtarış texnologiyasını inkişaf etdirərək, WAP, i-mode, J-SKY və EZWeb rejimləri ilə yanaşı AT&T Wireless, Sprint PCS, Nextel, Palm, Handspring və Vodafone şirkətləri də daxil olmaqla simsiz texnologiyalar təklif edən aparıcı şirkətlərə öz xidmətlərini təqdim edir.

2006-cı ildə Google şirkəti videoçarxların mübadiləsi sistemi olan Youtube saytını aldı. Bu gün milyonlarla videoçarxı özündə saxlayan Youtube saytına hər dəqiqə ərzində 60 saatlıq video yüklənir. 2007-ci ildə mobil qurğular üçün Android əməliyyat sistemi, 2008-ci ildə Google Chrome brauzeri istifadəyə verildi, 2011-ci ildə isə Google+ layihəsi həyata keçirildi.

Hal-hazırədək Google Android əməliyyat sistemi, Youtube, Google Plus, Nexus qurğuları, axtarış və nizamlaşdırma mexanizmlərinin təkmilləşdirilməsi və digər sahələrdə müxtəlif yeniliklər və yeni texnologiyalar tətbiq etməkdə davam edir. Hazırda Google tərcüməçi sistemində dillərin ümumi sayı 80-i ötüb keçmişdir.

9.4. Google sisteminin axtarış alqoritmləri

Google sistemində informasiyanın axtarılması və nəticələrdən ibarət səhifənin yaradılması prosesi özündə bir çox komponentləri birləşdirir. Google axtarışın dəqiqliyini yüksəltmək üçün daim özünün axtarış texnologiyalarını

təkmilləşdirir, biliklər şəbəkəsi və canlı axtarış kimi unikal innovativ həllərdən istifadə edir və layihələr həyata keçirir.

Google axtarış sistemində istifadə olunan proseslərin müxtəlifliyi barədə təsəvvürə malik olmağa imkan verən layihələr aşağıda baxılır.

9.4.1. İnternet saytların indeksləşdirilməsi alqoritmi

Google axtarış robotu Googlebot adlanan UserAgent moduluna malikdir. Googlebot axtarış indeksinin yaradılması üçün web-səhifələrin məzmununu skanlaşdıran əsas robot proqramıdır. Google sistemində bu robot proqramından əlavə daha bir neçə xüsusişdirilmiş robotlar mövcuddur:

- *Googlebot-Mobile* – mobil qurğular üçün web-saytları indeksləşdirən robot;
- *Google Search Appliance (Google) gsa-crawler* – yeni Search Appliance aparat-proqram kompleksinin axtarış robotu;
- *Googlebot-Image* – şəkillərin indeksləşdirilməsi üçün səhifələri skanlaşdıran robot;
- *Mediapartners-Google* – AdSense məzmunu müəyyənləşdirmək üçün səhifənin məzmununu skanlaşdıran robot;
- *Adsbot-Google* – AdWords məqsədli səhifələrin keyfiyyətinin qiymətləndirilməsi üçün məzmunu skanlaşdıran robot.

Yeni saytların indeksləşdirilməsi zamanı onlara axtarış göstəricilərinin ilkin qiymətlərinin mənimsədilməsi üçün Google tərəfindən iki yanaşmadan istifadə olunur. Birinci yanaşmaya əsasən, yeni domen adlarına malik olan, sahiblərini və server adlarını tez-tez dəyişən saytlar “gözləmə zonasına” yerləşdirilir. Yalnız Google axtarış mexanizm-

ləri həmin saytları axtarış üçün hazır etdikdə, onlar həmin zonadan çıxarılır.

İkinci yanaşma birincinin əksinə olaraq, bəzi aspektlər (məsələn, avtoritetə malik saytlardan müraciət, aktual və unikal məzmun və s.) nəzərə alınaraq, yeni yaranmış sayta indeksləşdirmə zamanı birbaşa yüksək PageRank qiyməti mənimsədilir və onlar axtarışın nəticəsində yaxşı yerləri tuturlar. Lakin müəyyən vaxt keçdikdən sonra bu göstəricilər həmin saytlardan götürülür və onlara real göstəricilər mənimsədilir.

9.4.2. Sıralama (nizamlama) alqoritmi

Google sisteminin müasir alqoritmləri istifadəçilərin nə axtarıqlarını dəqiq müəyyənləşdirmək və axtarış nəticəsində tapılmış sənədlərin onlara təqdim olunması ardıcılığını müəyyən etmək üçün 200-dən artıq müxtəlif signal və açardan istifadə edir. Bu zaman sözlərin səhifələrdə mövcud olması, səhifədə yerləşdirilmiş informasiyanın aktuallığı, istifadəçinin yerləşdiyi yer, səhifənin malik olduğu PageRank qiyməti və s. parametrlər nəzərə alınır.

PageRank – istinadlara görə nizamlama alqoritmidir. Bu alqoritm bir-biri ilə hiperistinadlarla bağlı olan sənədlər (web-səhifələr) kolleksiyasına tətbiq olunur. Burada hər bir sənədə onun digər sənədlərlə müqayisədə “vacibliyini” və ya “avtoritetliyini” göstərən müəyyən ədədi qiymət mənimsədilir. Bu ədədi qiymət web-səhifəyə istinadların sayından və həmin istinadların məxsus olduğu web-səhifələrin səviyyəsindən asılı olur. Belə ki, hər hansı web-səhifəyə istinadların sayı nə qədər çox olarsa, o, bir o qədər “vacib” hesab olunur. Eyni zamanda, web-səhifəyə istinad edən digər web-səhifələrin çəkilişi (“vaciblikləri”)

nə qədər yüksək olarsa, istinad olunan səhifənin “vacibliyi” də bir o qədər yüksək olar.

Başqa sözlə, PageRank – web-səhifəyə başqa web-səhifələrdən istinadların vacibliyini nəzərə almaqla onun çəkisinin hesablanması üsuludur. Burada PageRank qiyməti (vaciblik dərəcəsi) hesablanan web-səhifə *akseptor-səhifə*, bu web-səhifəyə istinad edən digər web-səhifə isə *donor-səhifə* adlanır. PageRank axtarışın nəticələrində saytların nizamlanması zamanı istifadə olunan köməkçi amillərdən biri olub, sorğuya görə tapılmış saytların axtarışın nəticəsi kimi istifadəçiyə təqdim olunması ardıcılığını müəyyən etmək üçün Google tərəfindən istifadə olunur.

S.Brin və L.Peyc tərəfindən təklif olunmuş PageRank üsulunda web-səhifələrin vaciblik dərəcəsinə müəyyən etmək üçün aşağıdakı düsturdan istifadə edilir:

$$PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + \dots + \frac{PR(T_n)}{C(T_n)} \right),$$

Burada PR – web-səhifəyə mənimsənilən PageRank qiymət, d – donor-səhifənin akseptor-səhifəyə ötürdüyü çəki hissəsini əks etdirən əmsal (adətən onun qiymətini 0,85 götürürlər, yəni donor-səhifə akseptor-səhifələrə çəkisinin 85%-ə qədərini verə bilər), n – web-səhifəyə (akseptor-səhifəyə) istinad edən web-səhifələrin (donor-səhifələrin) sayı, T_i – i -ci istinad edən web-səhifə, C – donor-səhifədə olan kənar istinadların sayıdır.

9.4.3. Biliklər qrafı (Knowledge Graph)

Google tərəfindən reallaşdırılan bu texnologiyanın əsasını semantik axtarış təşkil edir. Biliklər qrafı istifadəçinin nə istədiyini daha dəqiq başa düşməyə, axtarış sahəsini

(zonasını) müəyyən etməyə və axtarışın dəqiqliyini artırmağa imkan verir. Biliklər qrafının tətbiqi aşağıdakı üstünlükləri yaradır:

- axtarış sorğusunun məqsədini daha yaxşı başa düşmək;
- axtarış sorğusuna uyğun olan axtarış sahəsini (zonasını) genişləndirmək;
- axtarış sorğusuna cavab olaraq informasiyanın verilməsini sadələşdirmək (əgər üç ayrı-ayrı sorğu mahiyyət etibarını ilə eyni olarsa, Google onlara eyni cavab verəcəkdir);
- axtarış ifadəsinin semantik təhlili hesabına daha dəqiq axtarış nəticəsini təklif etmək;
- kontekst reklam üçün daha yaxşı axtarış nəticəsi təqdim etmək.

9.4.4. Google hummingbird (kolibri) alqoritmi

Google sistemi axtarış sorğusuna daha geniş cavab vermək, nəticəni genişləndirmək və daha münasib variantın seçilməsinə imkan yaratmaq məqsədilə sinonim sözlərdən istifadə edir. Lakin bu zaman nəzərə alınır ki, sinonim sözlər kontekstdən asılı olaraq bəzən sinonim mənalarda çıxış etmirlər. Ona görə də, daha münasib cavabların seçilməsi üçün sözlərin semantik (mənaya görə) təhlili prinsipindən istifadə edilir. Bu isə Google sisteminə axtarış ifadələrini daha yaxşı və tez başa düşməyə imkan verir.

9.4.5. Google sisteminin web-axtarış spamları ilə mübarizəsi

Saytların indeksləşdirilməsi zamanı axtarış sistemləri tərəfindən yalnız saytın relevantlığı, nüfuzluluğu, etibarlılığı

və digər keyfiyyət göstəriciləri deyil, eləcə də spamlılıq dərəcəsi nəzərə alınır. İnformasiya axtarışının təşkili üzrə Google şirkətində müxtəlif qruplar fəaliyyət göstərir. Bu qruplardan biri axtarışın və onun nəticəsinin keyfiyyətinin təmin olunması ilə məşğul olur. Bu qrup tərəfindən 2011-ci ildə işlənib hazırlanmış “Panda” alqoritmi axtarışın keyfiyyətinin yüksəldilməsinə xidmət edir.

Digər bir qrup axtarış spamları ilə mübarizə üzrə fəaliyyət göstərir. 2012-ci ildə bu qrup tərəfindən hazırlanmış “Penguin” alqoritmi web-spamlarla mübarizə məqsədilə yaradılmışdır.

Hər gün internetdə spamları olan milyonlarla web-səhifə peyda olur. Google-da bu spamların istifadəçilərə mane olmaması üçün saytların avtomatik və əllə yoxlanması üçün effektiv üsullardan istifadə edilir. Özündə spam saxlayan web-səhifələr fırıldaqçılıq yolu (açar sözlərin süni şəkildə çoxlu sayda təkrarlanması, yüksək PageRank və rən istinadların alınması, görünməz, gizli mətnlərin istifadəsi və s.) ilə yüksək reyting qazanmağa və axtarışların nəticələrində yuxarı yerlərə (sətirlərə) qalxmağa çalışırlar. Nəticədə həqiqi faydalı saytlar axtarışın nəticəsi siyahısının sonuna doğru sıxışdırılır, lazımi məlumatların tapılması çətinləşir və axtarışın effektivliyini aşağı düşür.

Web-saytlarda müxtəlif növ web-spamlardan istifadə olunur. Lakin Google-un alqoritmləri web-spamların əksər növlərini identifikasiya etməyə (tanımağa) imkan verir. Qalan (avtomatik aşkar olunmayan) web-spamları Google-un əməkdaşları “əllə” müəyyənləşdirirlər. Google web-spam aşkar olunmuş saytların reytingini avtomatik olaraq aşağı salır. Web-səhifələrdən web-spamların təmizlənməsi üçün Google müxtəlif alqoritmlərdən istifadə edir. Onlardan əsas ikisi Google Panda və Google Penguin alqoritmləridir.

9.4.6. Google Panda

Yuxarıda qeyd olunduğu kimi, 2011-ci ildə Google axtarış sistemi Panda adlanan yeni nizamlama alqoritmi istifadəyə verdi. Bu alqoritm insan amilinə (davranış amilinə) yönəlmiş olması ilə əvvəlki işləmələrdən kəskin fərqlənirdi. Panda alqoritmi nizamlama zamanı digər adi meyarlarla yanaşı, istifadəçilərin davranışlarını ciddi nəzərə alır. Məsələn, yalnız istifadəçinin saytda olması vaxtı deyil, həmçinin onun saytdakı hərəkətləri (nə ilə məşğul olması), sorğunu təşkil edən elementlərə aid olan istinadlar üzrə keçidlər və s. nəzərə alınır. Nəticədə alqoritm bir çox saytları keyfiyyətli saytlara qoyulan tələblərə cavab vermədiyi üçün axtarışın nəticəsində sona doğru sürüşdürülür.

Saytların Panda alqoritminin süzgəcindən keçməməsinə səbəb ola biləcək əsas amillər aşağıdakılardır:

- kontentin (məzmunun) unikalıq səviyyəsinin aşağı olması;
- başlığına relevant olmayan çoxlu sayda web-səhifələrin olması;
- açar sözlərə görə həddən artıq optimallaşdırma (açar sözləri süni şəkildə istifadə etmə);
- həm ayrı-ayrı web-səhifələrə, həm də bütövlükdə sayta baxmaqdan imtina edilməsi faizinin yüksək olması;
- istifadəçilərin saytda keçirdiyi vaxtın orta qiymətinin aşağı olması;
- qeyri-relevant reklam;
- unikal və qeyri-unikal mətnlərin nisbətinin kəskin mənfi olması;

- məzmunun bir şəkildə (yeknəsəq) formatlaşdırılması və strukturlaşdırılması;
- daxili istinadların dar çərçivəli olması.

9.4.7. Google Penguin (Pinqvin)

2012-ci ildə Google Penguin alqoritmi fəaliyyətə başladıqdan sonra bir çox saytların axtarıqların nəticələrində mövqeləri aşağı düşməyə başladı. Bu alqoritm də digər alqoritmlər kimi süzgəc funksiyasını yerinə yetirir. Onun qarşısında qoyulmuş vəzifə axtarış sisteminin keyfiyyətsiz informasiya ehtiyatlarından təmizlənməsindən ibarətdir.

Google Penguin alqoritmi aşağıdakı səbəblərdən saytların reytingini aşağı sala və ya onları axtarışın nəticəsindən çıxara bilər:

- bir-birinə istinad çəkilməyi ötürmək məqsədilə digər tərəfdaş saytlarla fəal istinad əlaqələrinin qurulması və axtarış alqoritmlərinin aldadılması;
- saytdan çıxan istinadların satılması, başqa sözlə saytdan digər sayta qeyri-tematik istinadların qoyulması;
- axtarış trafikini cəlb etmək və reklam vasitəsilə qazanmaq məqsədilə böyük sayda şablon səhifələrin istifadəsi;
- açar sözlərin daxili və xarici istinadlar üçün həddən artıq istifadəsi;
- digər internet ehtiyatlardan materialların istifadəsi;
- kataloq və bloqlarda keyfiyyətsiz məqalələr vasitəsilə yaradılmış web-spamların mövcudluğu;
- mətninin və meta-teqlərin yaradılması zamanı çoxlu sayda açar sözlərin süni istifadəi yolu ilə web-səhifənin saytda lüzumsuz optimallaşdırılması;

- reklamlarla həddən artıq yüklənməsi səbəbindən əsas səhifənin axtarışın nəticəsində mövqeyini itirməsi.

9.4.8. Google axtarış sisteminin əsas məhsulları

Google AdWords proqramı. Google AdWords proqramı reklamçılara öz məhsul və xidmətlərinin maraqlı (bu məhsul və xidmətlərlə maraqlana biləcək) şəxslərə çatdırılması imkanını verən global reklam proqramıdır.

Google AdSense proqramı. Google AdSense proqramı Google Network məzmunlu saytlar şəbəkəsinə daxil edilmiş web-saytlara onların məzmunu ilə sıx bağlı olan qrafik və mətn tipli reklam və elanları avtomatik göstərməyə imkan verir. Bu reklam və elanlar elə seçilir ki, onlar həqiqətən də istifadəçiləri (oxucuları) kifayət qədər maraqlandırmış olsun.

Google bildiriş. Google sistemində istifadəçinin verdiyi axtarış meyarlarına uyğun olan yeni axtarış nəticələri meydana çıxdıqda ona elektron poçtu vasitəsilə avtomatik olaraq məktub göndərilir. Hazırda üç növ bildiriş təklif olunur:

- Xəbərlər üçün bildiriş;
- Web-axtarış üçün bildiriş;
- Xəbərlər və web-axtarış üçün bildiriş.

Xəbərlər üçün bildiriş elektron poçtu vasitəsilə göndərilən məlumat olub, Google News üzrə axtarışın nəticəsində birinci onluqda yer alan yeni məqalələr barədə xəbərdarlıq edir. Web-axtarış üçün bildiriş elektron poçtu vasitəsilə göndərilən məlumat olub, istifadəçinin daxil etdiyi axtarış sorğusu üzrə nəticədə ilk iyirmi sırada yer alan yeni web-səhifələrin peyda olması haqqında xəbər verir. Xəbərlər və web-axtarış üçün bildiriş əvvəlki iki bildirişi özündə

birləşdirir və istifadəçinin sorğusuna uyğun olaraq aparılan axtarış nəticəsində həm Google News üzrə ilk onluqda, həm də web-axtarış üzrə ilk iyirmilikdə olan yeni məqalələr barədə məlumat verir.

Bloqlar üzrə Google axtarış. Bloqlar üzrə Google axtarış internetdə olan bloqların məzmunları üzrə ən sadə axtarış növüdür. Google-un web-axtarış sistemi tərəfindən idarə olunan bloqlar üzrə axtarış modulu şəbəkəyə qoşulmuş olan milyonlarla bloqlar (bloqların adları, müəllifləri, tarixləri və məlumatları) üzrə axtarış aparır və yeni, relevant nəticəni təqdim edir.

Google kitab axtarışı. Google kitab axtarışı kitabların məzmunlarının indeksi olub, istifadəçiyə lazımı kitabı tapmağa imkan yaradır. Adi kataloqa müvafiq olaraq Google kitab axtarışı kitabı tam şəkildə mağazadan və kitabxanadan əldə etməyin mümkün olması barədə məlumat verir. Bu xidmət istifadəçiyə dünyanın müxtəlif mağaza, kitabxana və nəşriyyatlarında olan milyonlarla kitabların mətnləri üzrə axtarış aparmağa, kitabların qısa məzmununa qabaqcadan baxmağa, onları elektron və ya çap məhsulu şəklində əldə etməyə kömək edir.

Google Catalogs. Google Catalogs vasitəsilə 6600-dən artıq bağlama ticarəti kataloqlarının tam mətninə giriş təqdim edilir.

Google-un istifadəçi axtarışı sistemi. Google-un baza axtarış texnologiyasının tətbiqi ilə yaradılmış və sazlanan istifadəçi axtarış sistemi axtarışın nəticəsini istifadəçi tərəfindən göstərilmiş web-saytlar və web-səhifələr üzrə nizamlayır və yaxud məhdudlaşdırır. Bu sistem axtarış prosesini spesifik məsələlərə və konkret bilik sahələrinə uyğun sazlamağa imkan verir.

Google-un kataloqu. Google kataloqu mövzu kateqoriyaları üzrə təşkil olunmuş web-saytlara baxmağa və onlar üzrə axtarış aparmağa imkan verir. Google-un web-kataloqu axtarış texnologiyasını Open Directory project kateqoriyalaşdırma sistemi ilə birləşdirir və dünyanın 75 dilində girişlə təmin edilir.

Dodgeball. Dodgeball mobil rabitə istifadəçilərin yerləşdiyi yerlər nəzərə alınmaqla qarşılıqlı fəaliyyətinin əlaqələndirilməsinə kömək edən şəbəkə xidmətidir.

Google-maliyyə. Google-maliyyə xidməti fondlar, açıq və qapalı aksioner şirkətlər haqqında işgüzar və maliyyə məlumatlarına tez və sadə girişi təqdim edir.

Gmail. Gmail Google-un pulsuz elektron poçt xidmətidir. Gmail istifadəçilərə pulsuz yaddaş və Google-un daxili axtarış texnologiyasını təklif edir. Gmail güclü süzgəc vasitəsilə daxil olan elektron poçt göndərişləri arasında spamları (arzu olunmaz poçt göndərişlərini) aşkar edir və istifadəçilər üçün yarana biləcək narahatçılığı aradan qaldırır. Eyni zamanda, Gmail antivirus qorumasını təmin edir.

Hazırda Gmail dünyanın əksər dillərində istifadə olunur. Gmail daxil olan poçtun avtomatik başqa ünvanə göndərilməsini və POP protokolu vasitəsilə daxilolmanı reallaşdırır. Onun xidmətləri internetə qoşulmuş mobil qurğuların (telefonların, planşetlərin və s.) istifadəçiləri üçün də açıqdır.

Google qrupları. Şəxsi göndərmə və xəbərdar etmə siyahılarının yaradılması, eləcə də açıq müzakirələrin təşkili üçün Google qruplardan istifadə etmək olar.

iGoogle. iGoogle istifadəçilərə Google saytının özləri tərəfindən yaradılmış əsas səhifəsinə ani girişi təmin edir.

İstifadəçilər bu səhifəyə Gmail və Google News xidmətlərini, konkret saytlara əlfəcini (qoşulma üçün nişanı), hava proqnozunu, kinofilmlərin cədvəlini və s. yerləşdirə bilərlər. iGoogle səhifəsi Google-un əsas səhifəsini əvəz etmir. Belə ki, istifadəçilər “Klassik səhifə” və ya “şəxsi səhifə” seçməklə iGoogle-un əsas səhifəsini Google-un əsas səhifəsi ilə və ya əksinə dəyişə bilərlər.

Google şəkillərin axtarılması. Google şəkillərin axtarılması indeksində internetdə nəşr olunmuş milyardlarla təsvirlər və onlar haqqında məlumatlar saxlanılır. Google şəkillərin axtarılması funksiyasını daha faydalı və əlverişli etmək üçün təsvirin ölçüsü, formatı, parlaqlığı və əlvanlığı, eləcə də verilmiş saytlar və domenlər üzrə geniş axtarış imkanları təklif olunur.

Google Labs. Google Labs Google şirkətinin mühəndisləri və novator istifadəçiləri üçün nəzərdə tutulmuş məşq poliqonudur. Google Lab-da yeni məhsulların prototipləri yerləşdirilir, həmin texnologiyanın tətbiqi və təkmilləşdirilməsi ilə bağlı istifadəçilərdən təkliflərin verilməsi xahiş olunur.

Google Local. Google Local xidməti istifadəçilərə poçt indeksləri, şəhər adları və konkret ünvanlar üzrə yerli təşkilatlar və müəssisələr haqqında relevant informasiyanı tapmaq imkanı verir.

Google Map. Google Map ərazinin xəritəsini təqdim edir. Onun köməyi ilə istifadəçi yer üzərində istənilən ərazinin xəritəsinə, şəhərin, küçələrin planına baxa, marşrut üzrə hərəkət istiqamətini müəyyən edə bilər. Bu zaman xəritəni sağa, sola, yuxarı və aşağı sürüşdürmək, xəritənin miqyasını artırmaqla təqdim olunan təsviri böyütmək mümkündür.

Mobil qurğular üçün Google. Mobil qurğular üçün Google məhsulları telefonlar, planşetlər və s. üçün xüsusi olaraq elə şəkildə işlənib hazırlanmışdır ki, zəruri informasiya həmin qurğuların ekranına yerləşə bilsin. Mobil qurğular üçün Google məhsulları aşağıdakı funksiyaları təqdim edir:

- web-axtarış;
- şəkillərin axtarışı;
- Google Map;
- WAP-axtarış;
- Google SMS (Google-un qısa məlumatlar xidməti).

Google News. Google News xidməti dünyanın 10000-dən artıq xəbər mənbəsindən məlumatın toplanmasını həyata keçirir, xəbər məlumatlarını internetdə dərc olunan sonra bir neçə dəqiqə ərzində axtarış üçün müvafiq formatda təqdim edir. Mühüm xəbərlər haqqında məlumatları Google News xidmətinin əsas səhifəsində başlıqlar şəklində yerləşdirir. Bu başlıqlar təsvir üçün heç bir siyasi baxış və ideologiyadan asılı olmadan yalnız kompüter alqoritmindən istifadə olunmaqla seçilir.

Mobil qurğular üçün fərdi əsas səhifə. iGoogle səhifəsinin mobil versiyası kompüter versiyasında olduğu kimi eyni məzmunu malik olur. Lakin mobil qurğular üçün fərdiləşdirilmiş əsas səhifə istifadəçiyə lazım olan bütün məzmunun xüsusi olaraq kiçik ekranlarda oxunması üçün rahat olan formatda özündə birləşdirir.

Fərdiləşdirilmiş axtarış. Google tərəfindən həyata keçirilən fərdiləşdirilmiş axtarış ayrıca istifadəçi üçün vacib olan nəticənin ona verilməsini təmin edir. Web-axtarışın tarixi funksiyası istifadəçilərə özlərinin apardıqları axtarışın tarixçəsini müşahidə və idarə etməyə, hansı nəticələr-

dən istifadə etdiklərini görməyə imkan verir. Vaxt keçdikcə bu məlumat fərdiləşdirilmiş axtarışın təşkili üçün istifadə olunur.

Malların axtarışı. Malların axtarışı istifadəçilər (alıcılar) tərəfindən internet vasitəsilə şəbəkədə və mağazalarda lazım olan malların axtarılıb tapılması və alınmasını həyata keçirmək üçün Google tərəfindən işlənib hazırlanmış axtarış mexanizmidir.

Google Reader. Google Reader web-texnologiya üzərində qurulmuş oxuma proqramı olub, internetdə lazımi məlumatların aşkarlanması və abunə yazılışının təşkilini asanlaşdırır. Bu funksiyadan istifadə etməklə istifadəçilər gündəlik istifadə olunan məlumatların internet vasitəsilə əldə olunmasını təşkil edə və beləliklə, son hadisələr barədə daim məlumat ala bilərlər.

Google Academy. Google Academy elmi ədəbiyyatın axtarılması üçün istifadə olunur. Burada elmi ədəbiyyat dedikdə geniş dairədə elmi nəşriyyatlar, peşəkar qruplar, signal nüsxələrinin verilənlər bazaları, universitetlər tərəfindən nəşr olunan elmi materiallar (ekspertlər tərəfindən attestasiya olunmuş işlər, dissertasiyalar, monoqrafiyalar, avtoreferatlar, signal nüsxələri, referatlar, elmi məqalələr, texniki hesabatlar və s.) nəzərdə tutulur.

Google Video. Google Video peşəkarlar və həvəskarlar tərəfindən yaradılmış və internetdə yerləşdirilmiş mümkün videotəsvirlərin axtarılıb tapılması üçün nəzərdə tutulmuşdur. Google Video çərçivəsində Google Video Store videomağazası yaradılmışdır. Burada istifadəçilər nəhəng televiziya şəbəkələri, sərbəst prodüserlər, kinofilm çəkənlər və ayrı-ayrı müəlliflər tərəfindən istehsal olunmuş vi-

deomaterialların böyük assortimentini ala və kirayə götürə bilərlər.

Web-masterlər üçün alətlər. Google-un web-masterlər üçün pulsuz təklif olunan alətləri web-saytları Google-robotlar üçün daha rahat edir. Bu alətlər çərçivəsində web-masterlərə təklif olunan Sitemaps funksiyası sayt haqqında geniş məlumatı Google axtarış sisteminə çatdırmaq məqsədilə saytın xəritəsini özündə saxlayan faylı bu sistemə ötürə bilir.

Google-un web-axtarış imkanları. İnternetin milyonlarla web-səhifələrinə geniş giriş imkanları ilə yanaşı Google-un web-axtarış xidmətinə şəbəkədə istifadəçilərə məhz axtardıqlarını tapmağa imkan verən xüsusi funksiyalar əlavə edilmişdir. Bunlara aşağıdakıları aid etmək olar:

- *Genişləndirilmiş axtarış funksiyası* – istifadəçilərə daha mürəkkəb (bul cəbrindən istifadə etməklə, dillərin, ölkələrin və ya web-saytların meyarları üzrə məhdudlaşdırmaqla) sorğular tərtib etməyə imkan verir;
- *Düzgün yazılışın yoxlanması* – sorğuda orfoqrafik və ya texniki səhvlərin olması ehtimal edildikdə alternativ axtarış şərti təklif olunur;
- *Web-səhifələrin tərcüməsi* – fransız, alman, italyan, portuqal, rus, ispan və digər dillərdə nəşr olunmuş web-səhifələri ingilis və əksinə tərcümə edir.
- *Birja qiymət təyin etməsi* – aksiyalar və pay fondları haqqında məlumatlara istinadları təqdim edir;
- *Küçələrin xəritələri* – küçələrin xəritələrinə istinadları təqdim edir və avtomobil surmə üçün istiqamət verir;

- *Kalkulyator* – riyazi məsələləri, o cümlədən sadə arifmetik hesablamaların aparılması, mürəkkəb riyazi və ya fiziki əmsalların hesablanması, ölçü vahidlərinin çevrilməsi məsələlərini həll edir;
- *Mənalərin müəyyən edilməsi* - Google tərəfindən əlifba göstəricisinə daxil edilmiş məzmun əsasında sözlərin və ya ifadələrin mənalərini (təriflərini) təqdim edir;
- *Telefon kitabçası* – ABŞ-da küçə ünvanlarını, müəssisələrin və yaşayış evlərinin telefon nömrələrini təqdim edir;
- *Nömrənin izlənməsi* – FedEx, UPS və USPS bağlamalarının izləmə nömrələrini, nəqliyyat vasitələrinin identifikasiya nömrələrini, məhsulların kodlarını, şəhərlərin telefon kodlarını, patentlərin nömrələrini, FAA hava gəmilərinin qeydiyyat nömrələrini və FCC avadanlıqlarının identifikasiya nömrələrini daxil etməklə tez axtarış aparmağa imkan verir.
- *Səyahətçilər üçün məlumat* – insanlara ABŞ-ın aviareysləri, aeroportlarında gecikmələr və hava şəraiti barədə son məlumatları əldə etməyə imkan verir
- *İstinadların keş yaddaşa yazılması* – indeksləşdirmə zamanı web-səhifələrin ani çəkilməmiş şəkillərini təqdim edir.

Google Analytics. Google Analytics xidməti reklamçılara və naşirlərə marketing kampaniyalarını daha effektiv aparmağa, web-saytlarını daha məhsuldar etməyə imkan verir

Google Blogger. Blogger web-texnologiya üzərində qurulmuş aparıcı nəşretmə aləti olub, insanlara şəbəkə jur-

nallarından və bloqlardan istifadə etməklə öz məlumatlarını ani olaraq internetdə nəşr etməyə imkan verir.

Google-təqvim. Google-təqvim web-texnologiya üzərində qurulmuş, istifadəçilər üçün vacib olan bütün təqvim məlumatlarını, o cümlədən dostlar və qruplar tərəfindən birgə istifadə olunan təqvimləri bir yerdə saxlamağa imkan verən pulsuz xidmətdir.

Google-kod. Google şirkətinə məxsus elmi-texniki işləmələrlə maraqlanan kənar mütəxəssislər bu sayta daxil olaraq proqram məhsullarının ilkin kodlarını və API xidmətlərin siyahılarını əldə edə bilərlər.

Google Desktop. Google Desktop proqramı istifadəçilərə öz kompüterlərində asanlıqla axtarış aparmaq imkanı verən pulsuz müştəri proqramıdır. Google desktop elektron poçtu, fayllar, musiqi, fotoşəkillər, çatlar, baxılmış web-səhifələr və s. üzrə tammətnli axtarış aparmaq imkan verir.

Google Earth. Google Earth istifadəçilərə internet vasitəsilə Yer kürəsinin rəqəmsal 3D-modelini təqdim edən yeganə proqramdır.

Mobil qurğular üçün xəritələr. Mobil qurğular üçün xəritə proqramı mobil qurğular vasitəsilə ərazinin xəritəsi üzrə marşrutu müəyyən etməyə, lazımi yeri tapmağa, ərazinin sputnik vasitəsilə çəkilmiş təsvirlərinə baxmağa imkan verir.

Picasa. Picasa ani zaman kəsiyində təsvirlərin internetdən tapılması, redaktə edilməsi, eləcə də istifadəçinin kompüterində olan təsvirlərin internetdə nəşr olunması üçün nəzərdə tutulmuş Google-un saytıdan pulsuz yüklənə bilən proqram təminatı olub

Alətlər paneli. Google alətlər paneli pulsuz yüklənir və istifadəçinin web-brauzerinin pəncərəsinə əlavə olunur.

Google Chrome. Google Chrome istifadəçilərə internetdə sadə, asan, sürətlə və təhlükəsiz işləməyə imkan verən web-brauzerdir. 2008-ci ildə istifadəyə verilməsinə baxmayaraq, internet istifadəçilərinin əksəriyyəti bu brauzerdən istifadə edir.

Youtube. Youtube video təsvirlər, video çarxlar, filmlər, kliplər və s. bankı olub, istifadəçilərə bu bankda olan video materiallar üzrə axtarış aparmağa, onlara baxmağa, onları öz kompüterinə yükləməyə, eləcə də öz videolarını Youtube-da nəşr etməyə imkan verən, multimediya imkanlarına malik nəhəng sistemidir.

9.5. Google sisteminin axtarış imkanları

Google PDF, RTF, PostScript, Microsoft, Word, Microsoft Excel, Microsoft PowerPoint və digər formatlı sənədlərdə axtarış aparmağa imkan verir. Axtarış alqoritminin effektivliyini yüksəltmək və nəticənin keyfiyyətini daha da yaxşılaşdırmaq məqsədilə həyata keçirilmiş layihələr Google sistemində aşağıdakı imkanların reallaşdırılmasına gətirib çıxarmışdır:

- *Cavabların və ilkin zəruri məlumatların təqdim edilməsi.* Axtarışın nəticəsinin, hava haqqında sorğular üzrə cavabların, son hadisə və yarışların nəticələri, ayrı-ayrı faktlar barədə məlumatların dərhal istifadəçiyə təqdim edilməsi təmin olunur.
- *Sorğuların mənasının başa düşülməsi və avtomatik daxil edilməsi.* İstifadəçi tərəfindən daxil edilmə prosesində axtarış sorğusunun qabaqcadan xəbər veril-

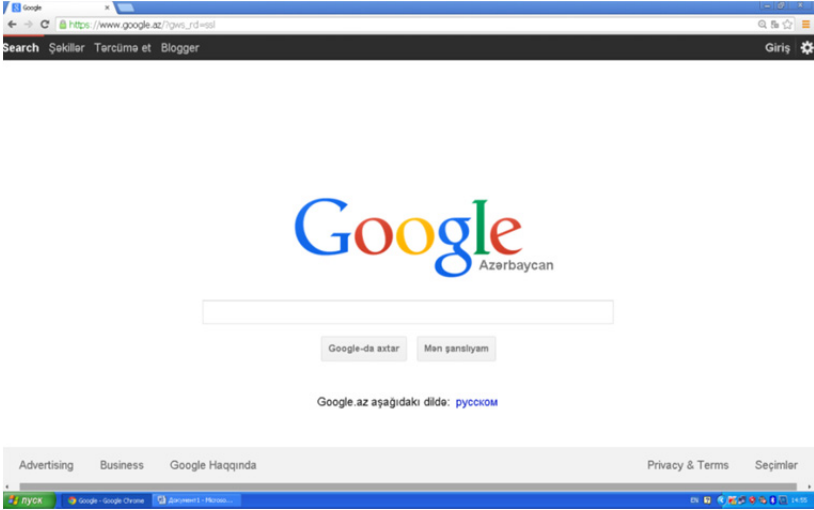
- məsi, daxil edilmiş simvollar ardıcılığına görə sorğunun başa düşülməsi və istifadəçiyə variantların təklif edilməsi, bir neçə mənası olan terminlərin tanınması, istifadəçi sorğularının mənasının başa düşülməsi reallaşdırılır.
- *Sinonimlər*. Oxşar mənalı sözlərin tanınması.
 - *Təqdim edilən məlumatların aktuallığı*. İstifadəçilərə ən son xəbərlər, məlumatlar və informasiya təqdim olunur, müəyyən tarixlər göstərməklə xəbər və məlumatların axtarılması həyata keçirilir.
 - *Canlı axtarış və səhifələrin təsviri*. Sorğunun daxil edilməsi zamanı axtarışın nəticəsi, tapılmış informasiya (məsələn, hər bir web-səhifənin başlığı və qısa təsviri) qabaqcadan istifadəçinin ekranına çıxarılır.
 - *Şəkillər*. Lazım olan səhifələrə keçidi dərhal təmin edən tapılmış şəkillərin kiçildilmiş ölçülü təsvirləri ekrana çıxarılır.
 - *Video*. Axtarışın nəticəsində lazımı videonu dərhal seçməyə imkan verən videoçarxların kiçildilmiş ölçülü təsvirləri təqdim olunur.
 - *İndeksləşdirmə*. Gələcəkdə axtarışın həyata keçirilməsi üçün internetdə olan təqdimat formasından asılı olmayaraq bütün növ sənədlərin toplanması və saxlanması həyata keçirilir.
 - *Biliklər şəbəkəsi*. Mövcud məlumat bazalarından real şəxslər, yerlər və əşyalar arasındakı əlaqələri nəzərə almaqla nəticələr çıxarılır və informasiya təqdim edilir.
 - *Mobil qurğular üçün xidmətlər*. Mobil qurğular (smartfonlar, planşetlər, netbuklar və s.) vasitəsilə in-

- ternetin xidmətlərindən istifadə edilməsi üçün xüsusi funksiyalar işlənilib hazırlanmışdır.
- *Dəqiqləşdirmə*. Axtarış sorğularını dəqiqləşdirməyə imkan verən genişləndirilmiş axtarışdan, oxşar sorğulardan və digər alətlərdən istifadə edilir.
 - *Təhlükəsiz axtarış*. Yalnız böyüklər üçün nəzərdə tutulmuş web-səhifələr, təsvirlər və videolar uşaqlar tərəfindən həyata keçirilən axtarışların nəticələrindən çıxarılır.
 - *Axtarış üsulları*. Şəkillər üzrə və ya səsli axtarış kimi yeni axtarış üsulları reallaşdırılmışdır.
 - *Saytların və səhifələrin keyfiyyəti*. Mənbənin etibarlılığını, mötəbərliyini və nüfuzluğunu müəyyən etmək üçün meyarlar toplusu istifadə edilir (məsələn, PageRank texnologiyası).
 - *Orfoqrafiya*. Mümkün orfoqrafik səhvlərin düzəldilməsi və alternativ variantların təklif edilməsi funksiyaları reallaşdırılmışdır.
 - *Tərcümə və beynəlmilləşdirmə*. Ölkələri və istifadəçilərin dilini nəzərə almaqla axtarış həyata keçirilir.
 - *Universal axtarış*. Axtarışın nəticəsinin ümumi vahid səhifəsində müxtəlif materiallar (şəkillər, xəbərlər, xəritə və video), eləcə də fərdi nəticələr təqdim edilir.
 - *İstifadəçinin konteksti*. Coğrafi yeri, axtarış tarixini və digər meyarları nəzərə almaqla daha dəqiq nəticələrin əldə olunmasını təmin edir.

9.6. Google axtarış sisteminin interfeysi

Google axtarış sisteminin əsas səhifəsi cox sadə görünüşə malik olub, heç bir şəkil və təsvirləri özündə saxlamır.

Google-un Azərbaycan variantı olan <http://www.google.az> saytının ümumi görünüş 9.3 sayılı şəkildə verilmişdir.



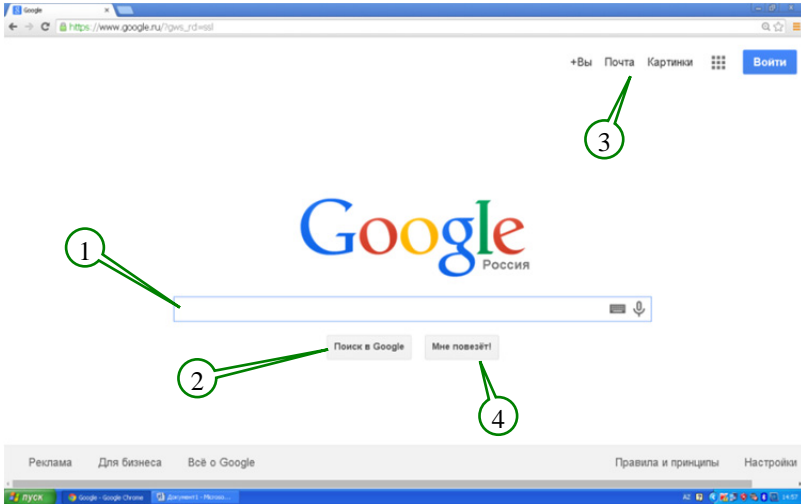
Şək. 9.3. Azərbaycan dilində Google axtarış sistemi

Azərbaycan variantında Google axtarış sisteminin funksiyaları tam əhatəli olmadığı üçün aşağıda şərhlər rus dilində olan <http://www.google.ru> saytı üçün verilir (şək.9.4). Google vasitəsilə internetdə informasiya axtarışını həyata keçirmək üçün axtarış sorğusunu sorğu sahəsində (1) daxil edib klaviaturada Enter və ya pəncərədə “Поиск в Интернете” (“Google-da axtar”) düyməsini (2) basmaq lazımdır.

Pəncərənin sağ yuxarı küncündə aşağıdakı funksiyalar təklif olunur (3):

- +*ВЫ* – Google+ xidmətinə keçid;
- *Почта* – elektron poçt xidmətinə keçid;
- *Картинки* – şəkillər üzrə axtarışın aparılması;
- *⋮* – digər funksiyalara keçid;

- *Войти* – Blogger rejiminə keçid.



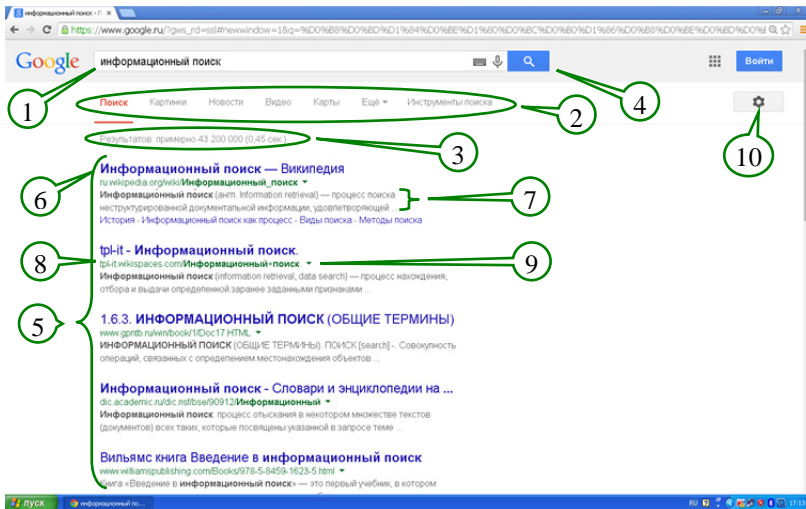
Şək. 9.4. Rus dilində Google axtarış sistemi

Google pəncərəsində axtarış sorğusunu daxil etdikdən və axtarış düyməsini basdıqdan sonra axtarışın nəticəsi pəncərəsi acılır (şək.9.5).

Pəncərədə axtarışın nəticəsi ilə yanaşı bir sıra məlumatlar yer alır. Axtarış pəncərəsində yer almış əsas elementlər aşağıdakılardır:

(1) – axtarış sorğusu sahəsi;
(2) – Google-un müxtəlif axtarış xidmətlərinə istinadları:

- *Поиск* – web-axtarış;
- *Картинки* – şəkillərin axtarılması;
- *Новости* – xəbərlər üzrə axtarış;
- *Видео* – videoların axtarılması;
- *Карты* – xəritə üzrə axtarış;
- *Еще* – digər xidmətlər;



Şək. 9.5. Google-un nəticə pəncərəsi (əvvəli)

Инструменты поиска – axtarış alətləri. Axtarış alətləri qismində dilin seçilməsi, tarixin qeyd olunması, nəticələrin diapazonunun göstərilməsi, axtarış yerinin seçilməsi təklif olunur.

(3) – axtrış nəticəsində tapılıb verilmiş sənədlərin sayı və axtarış üçün sərf olunmuş vaxt;

(4) – Google-un axtarış düyməsi;

(5) – axtarış nəticəsində Google tərəfindən tapılıb verilmiş sənədlərin siyahısı, onlara istinad və qısa məlumat;

(6) – tapılmış sənədin (web-səhifənin) adı;

(7) – tapılmış sənədin məzmununda sorğu şərtləri (yarımyağlı şrifflə qeyd olunur) rast gəlinən fraqment;

(8) – axtarış nəticəsində tapılmış sənədin web-ünvanı;

(9) – oxşar sənədlərin axtarılması;

(10) – axtarışın sazlanması modulu. Bura aşağıdakı daxildir:

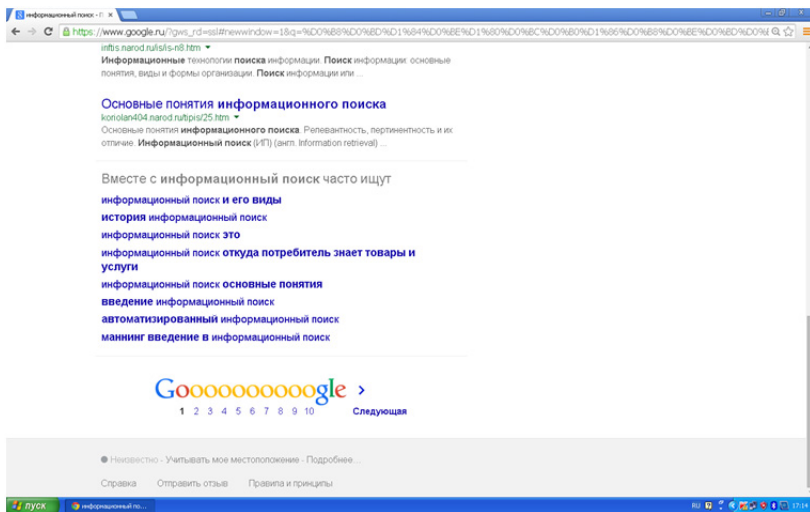
- *Настройка поиска* – axtarışın nəticəsinin təqdim olunması parametrlərinin sazlanması;
- *Языки (Languages)* – Google-da axtarış sisteminin və nəticələrin göstərilməsi dilinin seçilməsi;
- *Включить безопасный поиск* – təhlükəsiz axtarış süzğəcinin qoşulması (axtarışın nəticəsindən nalayiq məzmunların çıxarılmasını təmin edir);
- *Расширенный поиск* – Genişləndirilmiş axtarış (sorguda ayrı-ayrı sözlərin və söz birləşmələrinin daxil edilməsi və istisna olunması, məntiqi əməllərin, ədəd diapazonlarının, dilin, sözün rastgəlinmə yerinin, tarixin, formatın, domenlərin və s. seçilməsi imkanlarını təqdim edir);
- *История* – axtarış tarixçəsi ilə iş;
- *Поиск в справке* – Google axtarış sistemi barədə soruşca üzrə axtarış.

Google-un axtarış pəncərəsinin sonunda (şək.9.6) nəticənin təqdim olunması səhifələrinin ardıcılığını müəyyən edən “Google” sözünün dəyişmiş forması (“o” hərfinin sayı axtarışın nəticəsindən asılı olaraq çox olur) təqdim olunur. Növbəti axtarış səhifəsinə keçmək üçün növbəti “o” hərfinin və ya onun altında müvafiq hissənin səhifənin nömrəsini göstərən ədədin üzərində basmaq lazımdır.

9.7. Google-da axtarış sorgusunun tərtib edilməsi prinsipləri və şərtləri

Axtarış sorgusunun dəqiq tərtib edilməsi, axtarış şərtlərinin düzgün seçilməsi axtarışın uğurla nəticələnəcəyini təmin edir. Ümumi axtarışın aparılması üçün sorguda bir sözdən istifadə etmək olar. Məsələn, *Azərbaycan*. Lakin

effektiv nəticənin əldə edilməsi üçün sorğuda bir neçə axtarış sözündən istifadə edilməsi daha məqsədəuyğundur: ***Azərbaycan mehmanxana***. Qeyd etmək lazımdır ki, ***Azərbaycan mehmanxana qiymətlər*** daha yaxşı nəticə verə bilər.



Şək. 9.6. Google-un nəticə pəncərəsinin sonu

Google sistemində axtarış zamanı sorğuda hərflərin yuxarı və ya aşağı registrdə olması (böyük və ya kiçik hərflərlər yazılması) əhəmiyyət kəsb etmir. Belə ki, Google-da axtarış yalnız kiçik hərflərlə axtarılır. Ona görə də ***Lotfi Zadeh*** və ***lotfi zadeh*** eyni nəticəni verəcəkdir.

Ümumi halda, Google sorğuda əksini tapmış bütün sözlərin daxil olduğu sənədləri istifadəçiyə təqdim edir. Müəyyən edilmiş qaydalara görə, adi halda Google sorğuya daxil edilmiş sözlər arasında “and” məntiqi operatorunun olduğunu nəzərdə tutur, onu əlavə olaraq daxil etməyə zərurət yoxdur. Lakin sorğuya daxil edilmiş sözlərin ardıcılığı axtarışın nəticəsinə təsir edir. Bu baxımdan sorğunu

daha dəqiq formalaşdırmaq lazımdır. Məsələn, Qəbələ rayonunda istirahət barədə və ya Qəbələdə mehmanxana qiymətləri haqqında məlumat əldə etmək üçün ***qəbələdə istirahət*** və ya ***qəbələ mehmanxana qiymətləri*** sorğularını daxil etmək kifayətdir.

Qeyd olunmalıdır ki, axtarış prosesini ləngitdiyinə və nəticənin keyfiyyətinə müsbət təsir etmədiyinə görə Google ümumi sözləri, simvolları, tək rəqəmləri və hərfləri avtomatik olaraq sorğudan atır. Əgər nəticənin dəqiqliyi üçün hər hansı ümumi sözün sorğuda saxlanması zəruridirsə, onda onun qarşısına “+” işarəsi (əvvəlcə mütələq probel olmalıdır) qoyulmalıdır. Ümumi sözlərin sorğuda saxlanması üçün sorğunu tərtib edən söz birləşmələrini dırnaq işarəsi arasında da yazmaq olar. Məsələn, ***esperanto +1*** və ya ***“esperanto 1”***.

Google sözlərin hallarını nəzərə almağı bacarır. Yəni sorğuda sözlərin istifadə olunduğu forma ilə yanaşı, digər formalarda da variantları tapılır və istifadəçiyə təqdim edilir. Məsələn, ***ingiliscə danışmaq üçün kömək*** sorğusuna cavab olaraq, həmçinin ***ingilis dilində danışmaq, ingilis dilində danışmağa kömək*** və s. sorğuların da cavabı veriləcəkdir. Sözlün dəqiq formasının axtarılması üçün həmin söz dırnaq arasında yazılmalı və ya onun qarşısına “+” işarəsi qoyulmalıdır. Məsələn, ***ingilis dilində “danışmaq”*** sorğusu nəticədə məhz ***danışmaq*** sözü daxil olan sənədləri verəcəkdir.

Sorğuda olan sözlərin və ya söz birləşmələrinin dəqiq formasını özündə saxlayan sənədlərin tapılması və axtarışın nəticəsinə daxil edilməsi üçün sorğu dırnaq arasında yazılmalıdır. Məsələn, ***“ingilis dilində danışmaq kitabı”***.

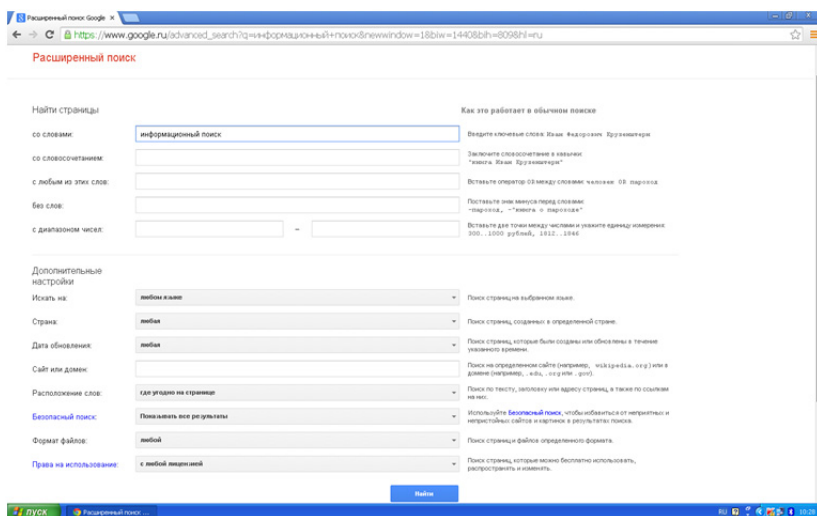
Bəzən sorğuya daxil edilmiş sözlər bir neçə mənə kəsb edir. Onların bəzilərini istisna etmək üçün sorğuya qarşısında “-“ işarəsi olmaqla istisnani göstərən sözü əlavə etmək lazımdır. Məsələn, internet şəbəkəsinə aid olmayan axtarış sistemləri barədə məlumat əldə etmək üçün *axtarış sistemi -internet* sorğusunu daxil etmək olar.

Google axtarış saytına daxil olarkən pəncərədə “Google-da axtarış” funksiyası ilə yanaşı “Mən şanslıyam” («Мне повезет») funksiyası (4) təklif olunur (şək. 9.4). Sorğunu daxil etdikdən sonra bu düyməni basdıqda Google tərəfindən sorğuya ən relevant səhifələr təqdim olunacaqdır. Ümumi halda, belə axtarış zamanı siyahı təklif olunmur, konkret sayta keçid baş verir. Əgər siyahı təklif olunarsa, onda siyahıda birinci yerdə “Mən şanslıyam” («Мне повезет») saytı olacaq. Məsələn, sorğu qismində *bakı dövlət universiteti* daxil edilərsə və “Mən şanslıyam” («Мне повезет») düyməsi basılırsa, onda axtarışın nəticəsi əvəzinə Bakı Dövlət Universitetinin saytı (bsu.edu.az) açılacaqdır.

Google axtarış sistemində daha dəqiq və faydalı axtarış nəticəsinin əldə edilməsi üçün *genişləndirilmiş axtarış rejimi* təqdim edilir. Şəkil 9.5-də göstərilmiş axtarışın sazlanması modulu (10) vasitəsilə açılan menyudan genişləndirilmiş axtarış (Расширенный поиск) bəndini seçməklə bu rejimə keçmək mümkündür. Bu zaman genişləndirilmiş axtarış pəncərəsi açılır (şək.9.7).

Genişləndirilmiş axtarış rejimində sözlərin, sinonimlərin və söz birləşmələrinin daxil olması, “OR” məntiqi əməliyyatının istifadəsi, ədədlər diapazonu, domenlər üzrə, eləcə də dillər, texnologiyalar (formatlar), tarixlər, sözün sənəddə rastgəlinmə yeri, domenlər, təhlükəsiz axtarış prinsipi nəzərə

alınmaqla axtarışın həyata keçirilməsini reallaşdıran operatorlar, funksiyalar və imkanlar təklif olunur.



Şək. 9.7. Google-un genişləndirilmiş axtarış rejimi

ƏDƏBİYYAT

Normativ-hüquqi aktlar

1. "İnformasiya, informasiyalaşdırma və informasiyanın qorunması haqqında" Azərbaycan Respublikasının Qanunu. Bakı, 3 aprel 1998-ci il.
2. "Elektron imza və elektron sənəd haqqında" Azərbaycan Respublikasının Qanunu. Bakı, 9 mart 2004-cü il.
3. "Dövlət sirri haqqında" AR. Qanunu. Bakı, 7 sentyabr 2004-cü il.
4. "Məlumat toplularının hüquqi qorunması haqqında" Azərbaycan Respublikasının Qanunu. 14 sentyabr 2004-cü il.
5. "İnformasiya azadlığı haqqında" Azərbaycan Respublikasının Qanunu. 01 yanvar 2005-ci il.
6. "İnformasiya əldə etmək haqqında" Azərbaycan Respublikasının Qanunu. 30 sentyabr 2005-ci il.
7. Azərbaycan dilinin izahlı lüğəti. 4 cildə. Bakı, 2006.

Tədris və elmi-metodiki vəsaitlər

1. Qasımov V.Ə. Elm və təhsilin informasiya təminatı sistemləri. Monoqrafiya. Bakı: Elm, 2005, 116 s.
2. Qasımov V.Ə. İnformatikanın əsasları. Dərs vəsaiti. MTN Akademiyası. Bakı, 2005. 86 s.
3. Abbasov Ə.M., Qasımov V.Ə., Quliyev R.A. İntellektual informasiya sistemlərində qərar qəbuletmə üsulları: Dərslik. Bakı, 2003, - 256 s.
4. Аббасов А.М., Касумов В.А. Интерфейсы пользователя в распределенных информационно-поисковых системах. Устройства систем и машин. 2003. №5. стр.67-74.

5. Байков В. Интернет. Поиск информации. Продвижение сайтов. Санкт-Петербург: БХВ, 2000. 288 с.
6. Гринберг И., Гарбер Л., Разработка новых технологий информационного поиска. // Открытие Системы. №9-10. 1999.
7. Гуде С.В., Ревин С.Б. Информационные системы. Учебное пособие. РЮИ МВД России. 2002. -149 с.
8. Дубинский А.Г. Модель мультиагентной системы информационного поиска в глобальной сети // Искусств. интеллект. 1999. №2. стр. 271-279.
9. Дубинский А.Г. Некоторые вопросы применения векторной модели представления документов в информационном поиске. // Управляющие системы и машины. 2001. №4. стр.77-83.
10. Дубинский А.Г. Факторы, влияющие на качество информационного поиска // Системний аналіз та інформаційні технології: Зб. тез доп. Міжн. наук.-практ. конф. студ., аспірантів та молод. вчених. - К.: НТУУ "КПІ", 2001. Ч.2. стр.43-48.
11. Ермаков А.Е. Тематический анализ текста с выявлением сверхфразовой структуры. // Информационные технологии. №11. Новосибирск. 2000. стр.37-40.
12. Ибушева О.В. и др. Интернет-технологии поиска и получения деловой информации. Тексты лекций. Нижнекамск. Нижнекамский химико-технологический институт. 2010. -42 с.
13. Касумов В.А. Методы информационного поиска в компьютерных сетях с сверхнасыщенными информационными ресурсами. Монография. Баку: Элм, 2004, 208 с.

14. Касумов В.А. Методы информационного поиска в Internet на основе нечетких отношений предпочтения. // Автоматика и Вычислительная Техника. Рига. Выпуск №4. 2003. стр.71-78.
15. Касумов В.А. Методы построения информационно-поисковых систем на базе иерархической модели информационного пространства Интернет. // Автоматика и Вычислительная Техника. Рига. Выпуск № 1. 2002. стр.40-51.
16. Касумов В.А. Моделирование информационного поиска в Интернет на базе нечетких знаний. // Автоматика и Вычислительная Техника. Рига. Выпуск: №2. 2002. стр.49-61.
17. Касумов В.А. Организация интерфейсов в поисковых системах. // Открытые системы. №9. Москва. 2001. стр.37. <http://www.osp.ru/os/2001/09/037.htm>
18. Касумов В.А. Организация системы поиска в Азербайджанской части Internet. // Открытые системы. №3. Москва. 2000. стр.59-62.
19. Козлов Д.Д. Информационно-поисковые системы в Internet: текущее состояние и пути развития. / Технологический обзор. - МГУ им. М.В. Ломоносова. Москва. 2000.
20. Корнеев В.В., Гарев А.Ф., Васютин С.В., Райх В.В. Базы данных: интеллектуальная обработка информации. - М.: Нолидж, 2000. 352 с.
21. Львович Я.Е., Леденева Т.М., Винокурова Т.Н. Лингвистическая модель информационно-поисковой системы. // Информационные технологии. №11. 2000. стр.22-31.

22. Панйр Дж. Вероятностные модели в информационно-поисковых системах. // "Nachr. Dok.". 1986. 37. №2. с.60-66.
23. Семенов Ю.А. Протоколы и ресурсы internet-М. Радио и связь, 1996. 320 с.
24. Солтон Дж. Динамические библиотечно-информационные системы. - М.: Мир. 1979. 558 с.
25. Храмцов П., Информационно-поисковые системы Internet. // Открытые Системы. №3. Москва. 1996.
26. Чечин Г.М., Положенцев Е.В., Нижникова С.В. "Поиск информации в сети Internet". Методические указания. - Ростов-на-Дону: РГУ, 2001. 41 с.
27. Попов А. Эффективная методика поиска информации в сети Интернет.//Citforum. <http://www.citforum.ru>.
28. Тихонов В. Поисковые системы в сети Интернет. «Поисковые системы в сети Интернет». <http://www/citform/ru/internet/search/starchsystems.shtml>.
29. Храмцов П. Моделирование и анализ работы информационно-поисковых систем Internet. // Открытые Системы. Москва. №6. 1996. <http://www.osp.ru/os/1996/06/46.htm>.
30. Manning Ch.D., Raghavan P., SchützeAn H. Introduction to Information Retrieval. Online edition. Cambridge. 2009. <http://www.informationretrieval.org/>.

İNTERNETİN ƏSAS AXTARIŞ SAYTLARI

1. <http://www.google.com> (<http://www.google.ru>,
<http://www.google.az>).
2. <http://www.bing.com>.
3. <http://www.baidu.com>.
4. <http://www.yahoo.com>.
5. <http://www.inktomi.com>.
6. <http://www.ask.com>.
7. <http://www.go.com>.(<http://www.infoseek.com>.)
8. <http://www.aport.com>
9. <http://www.webcrawler.com>.
10. <http://www.dmoz.com>.
11. <http://www.lycos.com>.
12. <http://www.hotbot.com>.
13. <http://search.aol.com>.
14. <http://www.yandex.ru>.
15. <http://www.rambler.ru>.
16. <http://www.citforum.ru>.
17. <http://www.osp.ru>.
18. <http://search-world.ru>.
19. <http://www.searchengines.ru>.
20. <http://www.searchengineland.com>.

Qasimov Vaqif Əlicavad oğlu

Texnika elmləri doktoru, professor

**İnformasiya axtarışı
üsulları və sistemləri**

Dərslik

Bakı – 2015

Redaktor:

Sevinc QOCAYEVA

Kompüter səhifələyicisi:

İsmayıl İMANOV

Yığılmağa verilib: 02.04.2015. Çapa imzalanıb: 10.06.2015

Fiziki çap vərəqi 18. Nəşrin formatı 60x84 ¹/₁₆.

Sifariş 545. Tiraj 200.

MTN-in Maddi-texniki Təminat Baş İdarəsinin
Nəşriyyat-Poliqrafiya Mərkəzində
nəşrə hazırlanmış və ofset üsulu ilə çap edilmişdir.