

ENSV INFORMATSIOONI INSTITUUT

E. Rannap

LEIUTISTE KIRJELDUSTE
ANALÜÜSI MEETODITEST

Kirjastus "Valgus"

Tallinn 1970

ENSV Informatsiooni Instituut

E. Rannap

Leiutiste kirjelduste analüüsi meetoditest

Kirjastus "Valgus"
Tallinn 1970

Arhiivkodu

6

R 15

Tartu Ülikooli Raamatukogu
ARHIIVKOGU

SISSEJUHATUS

Informatsioon uutest leiutistest omab esmajärgulist tähtsust tehnika edasiarendamisel, uute masinate, aparaatide loomisel.

Andmed uute leiutiste kohta avaldatakse patendikirjanduses (leiutiste kirjeldused, patendibülletäänid jm.).

Kogu maailmas väljaantud patentide juurde kuuluvate leiutiste kirjelduste hulk on käesoleval ajal ligikaudu 14 miljonit. Nendest kehtivate patentide osa moodustab umbes 23%.¹

Tööstusele vajalike leiutiste rakendamiseks, tehniliste lahenduste uudsuse kontrollimiseks jm. on vaja koostada süsteem, mis võimaldaks efektiivset informatsiooni otsingut.

Patendiinformatsiooni otsingusüsteemid kuuluvad dokumentalsete otsingusüsteemide hulka, s. t. otsingu tulemusena saadakse dokument, tema koopia või andmed dokumendi kohta.

Iga informatsiooni otsingusüsteemi efektiivsus sõltub mitmest tegurist, nagu informatsioonikeel, dokumendi analüüsi meetod, karakteristika koostamisel, tehnilised vahendid, mida kasutatakse informatsiooni salvestamiseks jm.

Olulisemad neist on informatsioonikeel ja dokumendi analüüsi meetod. Otsingu tulemus sõltub sellest, missugused andmed dokumendi kohta on valitud ja kuidas on dokumendi karakteristika (otsimistunnuste hulk) koostatud.

Käesolevas töös vaadeldakse mõningaid leiutise kirjelduse analüüsi võimalusi. Uurimus tehti Eesti NSV Teadusliku Tehnika-

¹ Прогноз развития научно-технической информации до 1980 г. М., 1969, с.96.

raamatukogu patendifondi baasil. Analüüsi peamiselt konstruktiivsete leiutiste kirjeldusi.

Töö esimeses osas iseloomustatakse leiutise kirjeldust kui juriidilise ja tehnilise informatsiooni allikat. Teises osas esitatakse ülevaade otsingusüsteemide koostamise meetoditest. Kolmas ja neljas osa käsitlevad leiutise kirjelduse ja vormeli analüüsi meetodeid. Töö lõpus analüüsitakse mõningaid tegureid, mis mõjutavad otsingusüsteemi efektiivsust.

1. LEIUTISE KIRJELDUS INFORMATSIOONIALLIKANA

Patendikirjanduses on kõige olulisem koht leiutiste kirjeldustel. Need on dokumendid, mis kuuluvad patentide või autoritunnistuste juurde. Nad sisaldavad juriidiliselt kaitstava leiutise sõnalise kirjelduse koos jooniste, skeemide ja mitmete õiguslike andmetega.

Leiutise kirjeldus peab täitma kolmenõuet:¹

- kirjeldama leiutist nii selgelt ja täielikult, et temas toodud andmete alusel oleks võimalik leiutist praktiliselt teostada;

- määrama lühidalt, selgelt ja täpselt leiutaja õiguste piirid, s.t. leiutise mahu;

- iseloomustama täpselt leiutise uudsust ja eesmärki.

Esimese nõude täidab leiutise kirjelduse täielik tekst, mille juurde kuuluvad vajaduse korral selgitavad joonised või skeemid. Teise ja kolmanda nõude täidab leiutise vormel. Leiutise vormel kujutab endast sõnaliselt antud tunnuste kogumit, mis määravad leiutise olemuse.

Espool toodud nõuded on esitatud NSV Liidu seadusandluse kohaselt koostatud leiutise kirjelduse kohta.

Patendiseaduste erinevuse tõttu erinevad ka paljude maade leiutiste kirjeldused nii vormilt kui sisult.

Leiutiste kirjeldused, mis antakse välja nn. saksa süsteemi kasutatavates riikides, sarnanevad ülesehituselt NSV Liidu leiutiste kirjeldustega. Hoolimata erinevustest, on leiutiste kirjelduste koostamisele esitatavad nõuded üldiselt ühesugused.

Leiutise kirjeldus on dokument, mis iseloomustab tehnika taset. Samal ajal on ta ka juriidiline dokument, mis määrab kindlaks autori õigused.

Leiutise kirjelduses toodud andmed võib seega jagada kahte rühma:

¹ Указания по методике экспертизы заявок на изобретения (33-2-67).
М., 1967, с.56.

- õiguslikku tähtsust omavad andmed;
- tehnilist sisu iseloomustavad andmed.

Toome järgnevalt olulisemad neist:¹

| Jrk. nr. | Leiutise kirjelduse element | Õigusliku iseloomuga andmed | Tehnilist sisu iseloomustavad andmed |
|----------|--|-----------------------------|--------------------------------------|
| 1 | 2 | 3 | 4 |
| 1. | Leiutise nimetus | + | + |
| 2. | Leiutise autor (tegelik leiutaja) | + | |
| 3. | Patendivaldaja | + | |
| 4. | Prioriteediandmed (riik, prioriteedikuupäev jm.) | + | |
| 5. | Patendi või autoritunnistuse number | + | |
| 6. | Klassifikatsiooniindeksid | | + |
| 7. | Patendi liik | + | |
| 8. | Leiutise kirjelduse tekstiline osa | | + |
| 9. | Joonised, skeemid | | + |
| 10. | Leiutise vormel | + | + |

Vastavalt leiutise kirjelduse iseärasustele on otsingusüsteemide loomisel mindud kahes suunas:

- süsteemid bibliograafiliste ja õiguslikku tähtsust omavate andmete ratsionaalseks salvestamiseks ja otsinguks;
- süsteemid leiutiste kirjelduste tehnilise sisu avamiseks.

Bibliograafilistel ja õiguslikel andmetel põhinevate süsteemide

¹ Täielikum patendikirjelduse elementide loetelu on toodud töös E. Rannap "Patendiinformatsioon", Tallinn, 1969.

mide väljatöötamine on küllalt lihtne, kuna kõik vajalikud andmed on leiutiste kirjeldustes olemas. Need andmed võimaldavad viia läbi järgmisi otsingu liike:

| Otsingu liik | Otsingu läbiviimiseks vajalikud andmed leiutise kirjeldusest |
|---|--|
| 1 | 2 |
| Temaatiline otsing | Klassifikatsiooniindeksid Leiutise pealkiri |
| Leiutaja või patendivaldaja patentide kindlaksmääramine | Leiutaja nimi Patendivaldaja nimi (firma nimetus) |
| Patendi kehtivuse määramine | Patendi number Patendi väljaandmise kuupäev |
| Paralleelpatentide (analooapatentide) otsing | Prioriteedi kuupäev, prioriteedimaa, esmase avalduse number, riik, kus esitati avaldus, patendi number, leiutaja nimi, patendivaldaja nimi |
| Numbriline otsing | Patendi (autoritunnistuse) number või ka avalduse number |

Siiani on palju keerukamaks osutunud niisuguste otsimisüsteemide loomine, kus oleks salvestatud küllaldaselt andmeid leiutise tehnilise sisu, tema olemuse kohta.

Käesoleval ajal on peaaegu kõikide riikide leiutiste kirjeldustel klassifikatsiooniindeks tema sisu iseloomustamiseks. Kogu maailmas on kasutusel ligikaudu 20 erinevat rahvuslikku patendiklassifikatsiooni. Klassifikatsioonide peamiseks puuduseks on nende hierarhiline ülesehitus. Iga klassi alajaotused on seda meelevaldsemad ja vähem usaldusväärsed tehnika edasise arengu peegeldamisel, mida kitsamateks üksikküsimusteks nad on jaotatud. Klassifikatsioonisüsteemide muutmise ja täiendamise nõuab paranduste sisseviiimist olemasoleva fondi leiutiste kirjeldustele. See nõuab suuri kulutusi.

Olukorda lihtsustab mõningal määral rahvusvahelise patendiklassifikatsiooni kasutuselevõtmine. Käesoleval ajal märgitakse

rahvusvahelise patendiklassifikatsiooni indeksid 24 riigi leiutiste kirjeldustel¹ (vt. lisa 1).

Eespool toodud puuduste tõttu on juba aastaid aktuaalne probleem, kuidas salvestada informatsiooni leiutiste kirjelduste kohta paljuhierarhiliste või mittehierarhiliste informatsioonikeelte abil.

2. OTSINGUSÜSTEEMIDE KOOSTAMINE

Otsingusüsteemi organiseerimisel arvestatakse üldiselt järgmisi tegureid:

- informatsioonivajaja kategooria;
- informatsiooni kasutamise eesmärk;
- informatsioonitöötajate hulk, kvalifikatsioon ja informatsiooniasutuse materiaalne varustatus ning rahalised võimalused.

Informatsioonivajajaid võib jaotada kategooriatesse mitmeti.

Näitena võime tuua järgmise jaotuse:

- fundamentaalteadustega tegelevad teadlased;
- rakendusteadustega tegelevad teadlased;
- konstruktorid, projekteerijad;
- mitmesugustel tootmisaladel töötavad spetsialistid.

Kui võtta aluseks ülaltoodud jaotus, siis võib näha, et Eesti NSV territoriaalse patendifondi kasutajate valdava enamiku moodustavad 3. ja 4. kategooria. Järelikult Eesti NSV-s kasutatakse patendiinformatsiooni esmajärjekorras tööstuses.

Tööstuse vajadusteks uuritakse patendimaterjale peamiselt järgmistel eesmärkidel:

- eeldatava leiutise uudsuse määramiseks (uudsuse eeleksper-tiis);
- seadme, masina, aparaadi patendipuhutuse määramiseks;
- uute tehniliste lahenduste rakendamiseks tööstuses.

¹ Вчеранний Р. П. и др. Патентная документация и патентная информация. М., ЦНИИПИ, 1969.

Patendiekspertiisi protsessis omab tähtsust teadmine, kas konkreetne ettepanek (tehniline lahendus) on juba tuntud (eel-ekspertiis) või kas tuntud tehniline lahendus on patendiga mingis konkreetnes välisriigis kaitstud.

Leiutiste otsingul nende rakendamiseks, juurutamiseks tööstuses on oluline teada, missugused antud probleemi tehnilise lahenduse variandid on üldse olemas ja missugused neist osutuvad hinnalisteks.

Nagu märgib R. Blühmstengel¹ (SDV), on praktiline töö näidanud, et otsingusüsteemi loomine ühe konkreetse tehnilise lahenduse leidmiseks on märksa lihtsam kui kõigi lahenduste väljaselgitamiseks.

See tuleneb muidugi asjaolust, et avaldaja kirjeldab oma konkreetset lahendust palju täpsemalt kui probleeme, mille lahendamiseks antud leiutis on sobiv. Samuti ei ole leiutise kirjelduses toodud kõiki leiutise rakendamise võimalusi.

Otsingusüsteemi loomisel tuleb otsustada, kuidas süsteem üles ehitada, kas valida ühekontuurne või mitmekontuurne süsteem. Ühekontuurse süsteemi puhul on dokument ja tema otsimiseks vajalikud ning piisavad andmed ühel ja samal informatsioonikandjal. Niisugust süsteemi kujutavad endast näiteks otsingusüsteemid mikrofilmidel, kus iga dokumendi juurde on salvestatud kodeeritult ka tema sisu iseloomustavad tunnused.

Oma olemuselt kuulub niisuguse süsteemi hulka ka patendikirjelduste fond, mis on korraldatud (paigutatud) klassifikatsiooniindeksite alusel.

Kahekontuursetes süsteemides on otsinguks vajalikud andmed dokumendi kohta eraldi nn. aktiivses mälu-seadmes. Otsingu tulemusena saadakse dokumentide aadressid, mille abil leitakse vajalikud dokumendid nn. passiivsest mälu-seadmest (hoidlast). Otsing viiakse läbi "aktiivses mälu-seadmes".

¹ Р. Блюмстенгель. Возможности международного сотрудничества в области патентной информации и патентной документации. Доклад на заседании секции "Проблемы патентно-информационного поиска". М., 1969, с.18. (Научный симпозиум "Изобретательство и научно-технический прогресс").

Niisuguse süsteemi realiseerimisel võib kasutada tavalisi kataloogikaarte, ääreperforatsiooniga kaarte, visuaalkaarte, ma-sinperfokaarte, elektronarvuti mäluseadmeid.

Mehhaniseeritud või automatiseeritud otsingusüsteemide loomisel on seni praktikas alati eelistatud kahekontuurset süsteemi, mis on eelistatum ka patendiinformatsiooni otsingusüsteemide juures. Sel juhul moodustavad "passiivse mäluseadme" leiutiste kirjelduste originaalid, nende ksero- või fotokoopiad, mikrofilmid.

"Aktiivse mäluseadme" organiseerimisel on vaja otsustada, mida tuleks salvestada informatsioonikandjale leiutise kirjelduse sisu iseloomustamiseks. Siin esineb mitu võimalust.

1. Leiutise sisu avamiseks kasutatakse leiutise kirjelduse peas leiduvaid andmeid (leiutise pealkiri, klassifikatsiooniindeksid). Sel juhul informatsioonikandja sisaldab bibliograafilise kirje koos klassifikatsiooniindeksitega.
2. Vahendusdokument (informatsioonikandja "aktiivses mäluseadmes") sisaldab bibliograafilise kirje, klassifikatsiooniindeksid, leiutise vormeli või referaadi (või ka annotatsiooni).
3. Vahendusdokument sisaldab bibliograafilise kirje ja märksõnade või ainerubriikidega laiendatud pealkirja.
4. Vahendusdokumendil on leiutise kirjeldust iseloomustavad märksõnad (need võivad esineda ka koos punkti 2 andmetega).
5. Vahendusdokument sisaldab ühe märksõna (deskriptori) koos kõigi seda märksõna sisaldavate dokumentide numbritega.

Otsingusüsteemi loomisel on võimalik rakendada kas otsest või inversioonimeetodit (sõnastikumeetod). Otsese meetodi puhul vastab igale algdokumendile (või tema koopiale) passiivses mäluseadmes vähemalt üks vahendusdokument aktiivses mäluseadmes. Vahendusdokument peab sisaldama küllaldaselt andmeid algdokumendi identifitseerimiseks.

Inversioonimeetodi rakendamisel salvestatakse informatsioonikandjale "aktiivses mäluseadmes" märksõna või deskriptor koos kõigi nende dokumentide aadressidega (kohaviitadega), mis sisaldavad seda märksõna (deskriptorit). Inversioonimeetodi kasutamine on võimalik ainult sel juhul, kui dokumendi analüüsi-

misel on tema sisu iseloomustamiseks valitud üksikud normaliseerimata märksõnad (unitermid), normaliseeritud märksõnad (deskriptorid) või ainerubriigid.

Otsingusüsteemi realiseerimise meetodist sõltub otseselt tehniliste vahendite (informatsioonikandjate) valik.

Otsese meetodi rakendamisel on võimalik kasutada tavalisi kataloogikaarte, ääreperforatsiooniga kaarte, pilukaarte, reitrikaarte, perfolinte, magnetlinte jm. Võimatu aga on kasutada näiteks visuaalkaarte, unitermkaarte.

Visuaalkaardid, unitermkaardid ja ka perfolindid, magnetlindid on sobivad inversioonimeetodi kasutamisel.

Järgmises tabelis on mõlemad meetodid vastandatud

märksõnad

| | d_1 | d_2 | ... | d_j | ... | d_m |
|-------|-------|-------|-----|-------|-----|-------|
| a_1 | 1 | 0 | ... | 1 | ... | 0 |
| a_2 | 0 | 1 | ... | 0 | ... | 1 |
| . | | | | | | |
| . | | | | | | |
| . | | | | | | |
| a_i | 1 | 0 | ... | 0 | ... | 1 |
| . | | | | | | |
| . | | | | | | |
| a_n | 1 | 1 | ... | 1 | ... | 0 |

$a_1, a_2, \dots, a_i, \dots, a_n$ - dokumentide numbrid

$d_1, d_2, \dots, d_j, \dots, d_m$ - märksõnad

tabeli iga rida vastab ühele dokumendile,
veerg - märksõnale (deskriptorile, ainerubriigile),
1 - tähendab märksõna ja dokumendi kokkukuuluvust,
0 - näitab märksõna ja dokumendi mittekokkukuuluvust.

Kui koostame vahendusdokumendid ridades antud andmetega, saame otsingusüsteemi otsesel meetodil, kui aga koostame vahendusdokumendi iga veeru kohta, saame inversioonimeetodil ülesehitatud süsteemi.

Informatsiooni otsingusüsteemi koostamisel tuleb arvestada ka tegureid, mis mõjutavad otsingusüsteemi efektiivsust, s.t. otsingul tekkivat informatsiooni kadu või ballasti.

Näiteks sünonüümide esinemine dokumentide karakteristikas võib põhjustada informatsiooni kadu, homonüümide esinemine toob kaasa otsingul ballasti.

Teguriteks, mis otseselt mõjutavad otsingu tulemusi, on informatsioonikeel, dokumendi karakteristikaga koostamise meetod ja otsingul kasutatavad tehnilised vahendid.

Järgnevalt vaatleme üht olulisemat tegurit, dokumendi karakteristikaga koostamist.

3. LEIUTISTE KIRJELDUSTE ANALÜÜS

Eesti NSV territoriaalse patendifondi baasil on läbi viidud uurimus otsingusüsteemide loomise võimalustest mittehierarhiliste informatsioonikeelte baasil. Selleks valiti kodumaiste leiutiste kirjelduste hulgast klassidesse 11c, 11d, 15a, 42m, 43a, 54d, 70c ja 75c kuuluvad leiutised teemal: "Informatsiooni salvestamiseks, säilitamiseks ja otsimiseks vajalikud seadmed".

Enamik vaadeldavate leiutiste kirjeldusi kuulusid nn. konstruktiivsete leiutiste hulka. Need on töövahendid, esemed jm.

Konstruktiivseid leiutisi võivad iseloomustada:¹

- mingite detailide (sõlmede) olemasolu või kasutamine seadmes;

¹ 33-2-67, c. 85.

- detailide (sõlmede) eriline paigutus üksteise suhtes või nende eriline kuju;
- detailide valmistamine mingist erilisest materjalist;
- detailide mõõtmete suhtelise suuruse muutmine.

Leiutiste iseloomustamiseks kasutati märksõnadena leiutiste kirjeldustes esinevaid loomuliku keele sõnu.

Leiutise olemuse kohta võib ettekujutuse saada:

- pealkirjast;
- klassifikatsiooniindeksitest;
- kirjeldusest;
- joonistest (või muudest graafilistest materjalidest);
- vormelist.

Leiutiste pealkirjade analüüs näitas, et selles on antud lühidalt leiutise üldnimetus. Näidatakse, mis liiki objektide hulka vaadeldav leiutis kuulub. Nimetus on alati esitatud ainsuses.

Näiteks: perforaator,
 lintperforaator,
 sorteerimismasin.

Pealkirjades ei ole antud tunnuseid, mis on vaadeldava leiutise korral uudsed.

Märksõnade hulk analüüsitud pealkirjades oli keskmiselt 6.

Leiutise kirjelduse analüüsimisel vaadeldi eraldi tema erinevaid osi: sissejuhatust, kirjeldavat osa ja vormelit. Nende erinevate osade iseloomustamiseks valiti märksõnad eraldi. Analüüsi tulemuste võrdlemine näitas, et kõigil vaadeldavail juhtudel kõik pealkirjas esinevad märksõnad korduvad vormeli esimeses pooles. Enamik sissejuhatuses esinevatest märksõnadest kordub täielikult ka vormelis. Kirjeldavas osas esines märksõnu tunduvalt rohkem kui teistes osades.

Leiutise kirjelduses on suur väärtus ka joonistel. Kirjeldav osa koos joonistega moodustab ühe terviku. Harva on kirjeldus joonisteta arusaadav. Spetsialistile piisab pilgu heitmisest joonistele, et mõista leiutise olemust. Ometi ei ole joonised sobivad märksõnade valikuks, kuna nad on ühetähenduslikud ainult koos tekstiga. NSV Liidus kehtivate leiutise kirjelduse koostamise

juhiste järgi on joonistel leiutise kirjelduse juures selgitav, illustreeriv funktsioon.

Kui analüüsimisel seada eesmärgiks nende märksõnade valimine, mis iseloomustaksid leiutise tehnilist olemust, siis piisab täielikult vormeli analüüsist. Vastavalt patendiseadustele (NSV Liidus "Juhised leiutise kirjelduse koostamiseks") peab vormel sisaldama lühikeses sõnalisel vormis kõik leiutise olulised tunnused.

Sel juhul puuduvad leiutise kirjelduse karakteristikast need märksõnad, mis näitavad leiutise kasutamise valdkonda ja ülesandeid.

Kui otsimissüsteem koostada nende riikide leiutiste kirjelduste kohta, kus on kasutusel nn. saksa klassifikatsioon, siis võib märksõnade valikul piirduda ainult vormeliga. Sel juhul iseloomustavad leiutise kirjeldusel vastava riigi patendiameti poolt märgitud klassifikatsiooniindeksid leiutise kasutamise valdkonda. Saksa klassifikatsioonisüsteemi kasutamisel määratakse liigitatavate objektide (leiutiste) kuuluvus mingisse patendiklassi tema kasutusala järgi. Näiteks:

Rotatsioonimasin perfokaartide valmistamiseks

Vastav klass 43 on "Kontrollseadmed ja kaubanduses kasutatavad automaadid".

Kui leiutis on kasutatav mitmes valdkonnas, märgitakse talle mitu klassifikatsiooniindeksit.

NSV Liidus kehtivate ekspertiisi juhiste järgi peab ekspert leiutise klassifikatsiooniindeksite määramisel kontrollima indeksite õigsust vormeli abil.¹

Seega vormelist valitud märksõnad koos klassifikatsiooniindeksiga peaksid andma küllaldase ettekujutuse nii leiutise olemusest kui ka kasutusala.

Esti NSV territoriaalsest patendifondist moodustavad põhilise osa nende riikide leiutiste kirjeldused, kus on kasutusel saksa klassifikatsioon.

Eespool toodule toetudes analüüsiti otsingusüsteemi loomisel

¹ 33-2-67, c. 25.

eksperimenti käigus ainult vormeleid, kusjuures vaadeldi analüüsil esinevaid erinevaid võimalusi.

4. LEIUTISE VORMELI ANALÜÜS

4.1. Märksõnade (unitermide) valik leiutise vormelist

Märksõnade valik dokumendi iseloomustamiseks põhineb oletusel, et dokumendi peamist sisu on võimalik küllaldase täpsuse ja täielikkusega väljendada teatud hulga märksõnadega, mis otseselt või kaudselt esinevad indekseeritavas tekstis. Selle idee esimene teostaja oli 50-ndatel aastatel Ameerika Ühendriikides Mortimer Taube.

Märksõnadeks nimetame antud juhul olulisi sõnu või sõnaühendeid, millel on tähenduslik funktsioon.

Vormelist on võimalik valida ja eraldada üksikud olulised sõnad. Nende valitud märksõnade korrastatud hulk koos bibliograafiliste andmetega moodustab leiutise karakteristikaks.

Tavaliselt valitud märksõnad esitatakse karakteristikaks alfabeetilises järjestuses. Märksõnad antakse kas ainsuse või mitmuse nimetavas käändes, hoolimata sellest, missugused grammatilises vormis nad esinesid analüüsitavas tekstis.

Näide. Autoritunnistuse nr. 189 242 vormeli piiravas osas on leiutise objektiks rotatsioonimasin perfokaartide valmistamiseks, mis koosneb trükkimissektsioonist, vastuvõtuseadmest, sõlmedest, mis võimaldavad pidevat perforeerimist ja lõikamist jne.

Sellest osast valitud märksõnad on järgmised: rotatsioonimasin, perfokaardid, valmistamine, trükkimine, perforeerimine, lõikamine.

Selle meetodi peamiseks puuduseks on sünonüümide esinemine. Sünonüümsuse vältimiseks võib hiljem kartoteeki sisse viia viidete süsteemi.

Üksikute märksõnadega on võimalik dokumenti iseloomustada ka sel juhul, kui me ei vali neid otseselt tekstist. Indekseerija võib leiutise kirjelduse karakteristikasse märkida sõnad, mis

esinevad analüüsitava tekstis kaudselt (tekstis esinevate üksikute märksõnade või nende rühmade sünonüümid).

Leiutise kirjelduse karakteristika koostamine üksikutest märksõnadest võib toimuda ka siis, kui juba enne analüüsimisele asumist on olemas vastava eriala märksõnastik. Sel juhul leiutise vormeli analüüsimisel valitakse dokumendi karakteristikasse märksõnastikust need, mis esinevad vormelis samal kujul või sünonüümidena.

Eespool kirjeldatud karakteristika koostamise meetodid võimaldavad otsingusüsteemi üles ehitada nii otsese kui ka inverse süsteemina. Et võimaldada otsingut märksõnade mitmesuguste kombinatsioonide alusel, on vajalik inversioonimeetodi rakendamise. Sel juhul "aktiivne mäluseade" sisaldab märksõnakartoteegi. Kaartidena on võimalik kasutada uniterm- või visuaalkaarte (vt. lisa 2 ja 3). Kaardi ülaosale kantakse märksõna, tabelisse (unitermkaardil) või ruudustikku (visuaalkaardil) leiutise kirjelduse järjekorranumber. Kui fond on korraldatud autoritunnistuste või patendinumbrate järgi, siis võib kasutada ka vastavat patendinumbrit. Fondi süstemaatilise paigutuse korral on vajalik "aktiivse mäluseadme" koostisena veel kolmas kartoteek, mis on koostatud otsese meetodiga. Selles vastab igale dokumendile üks kaart, millele on kantud dokumendi bibliograafiline kirje ja koha viit, mis näitab dokumendi asukohta hoidlas. Bibliograafilises kartoteegis märgitakse igale kaardile järjekorranumber vastavalt dokumentide registreerimisele (saabumisele fondi).

Märksõnade korrastatud hulk leiutise iseloomustamiseks, nende salvestamine märksõnakaartidele võimaldab otsimist läbi viia märksõnade mis tahes kombinatsiooni järgi. On ilmne, et see vaba kombinatsioonide moodustamise võimalus toob otsingul endaga kaasa ka palju ballasti. Kui meil on näiteks vaja leida leiutiste kirjeldused mikrofilmi kasutamisest visuaalkaartidena, on märksõnad vastavalt:

kasutamine, mikrofilm, visuaalkaart.

Kui me kartoteegist valime neid märksõnu sisaldavad kaardid ja teeme kindlaks, missugused dokumendid sisaldavad neid kolme märksõna, siis võime saada ballastina näiteks leiutiste kirjeldused, kus on juttu visuaalkaartide kasutamisest mikro-

filmide otsingusüsteemis. Me saame otsingu tulemusena kõik need dokumendid, kus antud märksõnad esinevad mis tahes järjekorras ja sellest hoolimata, missugused märksõnad neid dokumente veel ise-loomustavad.

Üldiselt, kui küsimus sisaldab näiteks n märksõna ja valekombinatsioonid tekivad kõige tõenäolisemalt k märksõnast, siis võime saada otsingul ballastina juba C_n^k dokumenti. See tähendab, et otsingusüsteem annab teatud kindla protsendi mitterelevantseid dokumente.

Valekombinatsioonide vältimiseks on oluline dokumendi karakteristikas ja ka informatsiooninõudes väljendada märksõnadevahelisi seoseid.

4.2. Süntagmaatilised seosed leiutise vormelis

Süntagmaatiliste seoste all mõeldakse a posteriori määratud lineaarseid sõnadevahelisi seoseid, mis ühendavad neid sõnarühmadeks ja lauseteks.¹

Loomulikus keeles kasutatakse süntagmaatiliste seoste väljendamiseks mitmesuguseid grammatilisi vahendeid, nagu:

- grammatiliste vormide moodustamist tüvele kindla tähendusega afiksrite liitmise teel;
- sidesõnu, abisõnu jm.;
- sõnade järjestust;
- jaotamist sõnarühmadeks (süntagmadeks) ja lauseteks.

Nendest grammatilistest vahenditest sobivad leiutise vormeli analüüsimisel sõnade järjestus ja jagunemine süntagmadeks. Teksti jaotust lauseteks ei saa kasutada, kuna üldreeglina leiutise vormel koosneb ühest lausest.

Vaatleme järgnevalt mõningaid süntagmaatiliste seoste väljendamise võimalusi leiutise vormelist dokumendi karakteristika koostamisel.

¹ А. М. Черный. Синтагматические отношения между дескрипторами. НТИ, 1968, сер. 2, № 4, с.6-16.

Dokumendi karakteristikat koostamine
märksõnade ahelatest

Dokumendi iseloomustamisel märksõnade ahelate abil valitakse tekstist need sõnarühmad, mis avavad kõige enam kirjutise peamist teemat. Indekseerija võib iseloomustamiseks valida ka niisugused fraasid, mis esinevad tekstis kaudselt.

Iga väljavalitud märksõnade rühm moodustab otsingusüsteemis iseseisva informatsioonikeele üksuse. Tavaliselt säilitatakse selles sõnarühmas ka grammatilised vormid.

Näitena vaatleme leiutise kirjelduse nr. 178 190 vormelli analüüsi.

Saame järgmised sõnarühmad.

Otsingusüsteem visuaalkaartidega.

Formaatse mikrofilmi kasutamine visuaalkaartidena.

Bibliograafiliste kaartide mikrokujutised otsingusüsteemis.

Need on esialgselt välja valitud ja leiutise olemust iseloomustavad sõnarühmad. On võimalik salvestada nad otsingusüsteemi niisugusel kujul nagu nad esinesid dokumendis. Sel juhul on nad ühetähenduslikud ja analüüsitava dokumendiga kõige enam seotud. Niisugusel korral on aga otsimine raskendatud, kuna ei ole tõenäoline, et küsimus sisaldaks dokumendi iseloomustamiseks valitud fraasidega identseid fraase. Järelikult tuleb märksõnade ahelad moodustada nii, et nad hõlbustaksid otsimist.

Üheks võimaluseks on järjestada märksõnad ühe sõnarühma piires põhimõttel "üldiselt üksikule", s.t. esimesele kohale paigutada kõige üldisem mõiste, edasi tema täpsustused ja omadused.

Näiteks:

otsingusüsteem - visuaalkaardid;

visuaalkaardid - formaatne mikrofilm;

bibliograafilised kaardid - mikrokujutis.

Otsingusüsteemis korraldatakse kaardid, millele on kantud märksõnade ahelad, esimese märksõna järgi alfabeetiliselt. Otsingu täielikkuse kindlustamiseks peaksid iga märksõnade ahe-

la kõik sõnad esinema üks kord esimese sõnana. See võimaldaks dokumenti leida kõigi ahelas leiduvate märksõnade järgi.

Kui üks märksõnade ahel sisaldab keskmiselt k elementi ja dokumendi kohta on koostatud m ahelat, siis andmed selle dokumendi kohta peaksid sisalduma $m \cdot k$ rubriigis. Kui aga k märksõnast koosnevast ahelast koostada veel täiendavalt rubriike, mis sisaldavad 1,2,3,..., k märksõna, saame ühest ahelast rubriike

$$\sum_{m=1}^k C_k^m = C_k^1 + C_k^2 + C_k^3 + \dots + C_k^k = 2^k - 1,$$

kus C_k^m on kombinatsioonide arv k elemendist m kaupa.

Vaadeldud juhul kindlustame otsingul täielikkuse, kuid otsingusüsteemi koostamine osutub väga tömahukaks ja tema realiseerimine kaartidel viib mammutkataloogi tekkimiseni.

Nende puuduste vältimiseks on välja töötatud mitmesuguseid meetodeid.

1964. aastal töötas Inglismaal J. P. Sharp välja märksõnade ahelate valiku meetodi SLIC (Selective Listing in Combination).¹

SLIC-meetodiga valikul on põhiprintsiibiks märksõnade järjestamine ahelas alfabeetiliselt. Ühest ahelast moodustatakse uued ühe-, kahe-, kolme- jne. kaupa, kusjuures alfabeetiline järjestus säilib. Meetodi selgitamiseks moodustame kombinatsioonid neljas terminist ühe-, kahe-, kolme- ja neljakaupa.

Olgu A, B, C, D terminid, siis kombinatsioonid on

A, B, C, D,

AB, AC, AD, BC, BD, CD,

ABC, ABD, ACD, BCD

ABCD

Nende kombinatsioonide analüüs näitab, et mõned neist moodustavad osa pikemast rubriigist. Nii näiteks A, AB, ABC kuuluvad

¹ J. P. Sharp. The SLIC Index. - "American Documentation", 1966, v. 17, No. 1, p. 41-44.

samas järjestuses ABCD-sse ja sellepärast ei ole oluline neid eraldi rubriikidena välja tuua. Samuti B, BC kuuluvad BCD-sse, C kuulub CD-sse ja AC ACD-sse. See annab võimaluse vähendada rubriikide arvu viieteistkümnelt kaheksale neljast märksõnast koosneva ahela puhul.

Üldine rubriikide arv k märksõnast koosneva ahela puhul on SLIC-meetodi rakendamisel $2^{(k-1)}$.

On ilmne, et

$$2^{(k-1)} < 2^k - 1.$$

Teine võimalus märksõnade ahelatest rubriikide moodustamiseks nii, et iga märksõna esineks üks kord esimesel kohal, on permutatiivregistrite koostamise meetodika rakendamine. Sel juhul kasutatakse otsingusüsteemi koostamiseks arvutusperforatsioonimasinaid või raale.

Seni on permutatiivregistreid koostatud peamiselt signaalinformatsiooni väljaandena (KWIC- ja KWOC-registrid)¹, kusjuures märksõnade ahel moodustatakse trükise pealkirjast. Märksõnade esinevad kõik terminid ja tähenduslikud sõnad, mis on olulised antud teaduse- või tehnikaharule. Märksõnad esinevad ahelas samas järjestuses nagu pealkirjaski. Iga märksõna esineb eraldi rubriigis, kusjuures nad antakse alati koos mikrokontekstiga. Kui ahel sisaldab k märksõna, siis kordub ta registris k korda.

Leiutist iseloomustava märksõnade ahela võime valida vormelist. Sealt võib ühe leiutise kohta moodustada mitu ahelat, võib valida ka ainult vormeli piirava osa (leiutise objekti koos eritunnustega).

Näide. Autoritunnistuse nr. 239 213 juurde kuuluva leiutise kirjelduse vormel:

Ääreperfokaartide säilitamise ja selekteerimise seade, mis koosneb korpusest ja selekteerimisvarrastest, erinev selle poolest, et seadme mõõtmete vähendamiseks ja käsitlemise mugavuse suurendamiseks on painduvad selekteerimisvardad paigutatud korpusesse monteeritud torukestesse.

¹ KWIC - Key-Word in Context;
KWOC - Key-Word out of Context.

Valime leiutise vormeli piiravast osast märksõnade ahela:
ääreperfokaardid, säilitamine, selekteerimine, seade, kor-
pus, selekteerimisvardad.

Leiutise kirjelduse nr. 279 939 vormel:

Perforeerimisseade, mis sisaldab veomehhanismi, elektromag-
neteid, puansoone ja matriitse, erinev selle poolest, et seadme
töökiiruse ja kindluse tõstmiseks on veomehhanism kujundatud võl-
likujulisena, mille peale on kinnitatud radiaalselt rullid või
kühmud, seostatuna löökpäästikuga ja puansooni tagastava lingiga,
mille profiilid on kujundatud vastavalt nõuetele, näiteks puansoo-
ni liikumiseks sinusoidaalse seaduse järgi.

Valime leiutise vormeli piiravast osast märksõnade ahela:
perforeerimisseade, veomehhanism, elektromagnetid, puansoo-
nid, matriitsid.

Moodustame nendest kahest ahelast permutatiivregistri näi-
te.

| Märksõnade ahelad | Leiutise nr. |
|--|--------------|
| <u>elektromagnetid</u> | |
| perforeerimisseade, veomehhanism, elektromag- netid, puansoonid, matriitsid | 279 939 |
| <u>korpus</u> | |
| ääreperfokaardid, säilitamine, selekteerimine, seade, korpus, selekteerimisvardad | 239 213 |
| <u>matriitsid</u> | |
| perforeerimisseade, veomehhanism, elektromag- netid, puansoonid, matriitsid | 279 939 |
| ----- | ----- |
| ----- | ----- |
| <u>ääreperfokaardid</u> | |
| ääreperfokaardid, säilitamine, selekteerimine, seade, korpus, selekteerimisvardad | 239 213 |

Märksõnade ahel leiutise 239 213 vormelist peab korduma ülaltoodud näites 6 korda, leiutise 279 939 märksõnade ahel 5 korda. Näites on märksõna toodud ahelast välja. Niisugusel kujul esitatud permutatiivregistrit nimetatakse KWOC-tüüpi registriks. KWIC-tüübi näiteks on referatiivajakirja "Научная и техническая информация" iga väljaande lõpus toodud permutatiivregister (vt. lisa 4).

Permutatiivregistritena koostatud otsingusüsteemi "aktiivse mälu" saame trükitud kataloogi kujul. Neid on võimalik teatud ajavahemiku järel täiendada ja koostada kumulatiivseid katalooge. Neid on võimalik paljundada mis tahes eksemplaride arvus.

Permutatiivregistrite peamiseks puuduseks on suur maht. Et aga nende koostamisel kasutatakse arvutusperforatsioonimasinaid ja elektronarvuteid, siis ei ole see takistuseks nende koostamisel.

Märksõnade osatähtsuse väljendamine dokumendi karakteristikas

Paljudes otsingusüsteemides kasutatakse otsingu tulemusena saadava ballasti vähendamiseks nn. märksõna osatähtsuse näitajaid. A. Tšernõi defineerib neid järgmiselt:

"Osatähtsuse näitajad - need on spetsiifilised sümbolid, mis seotakse deskriptoritega, et näidata neid funktsioone (loogilist osatähtsust), mida antud deskriptorid täidavad otsimiskarakteristikas või informatsiooninõudes. Osatähtsuse näitajad muudavad deskriptorite mõtet (nende abil väljendatud mõistete mahu vähendamise teel) ja meenutavad afikseid aglutineeritud keeltes."¹

Märksõnade osatähtsuse väljendamiseks võime üldistatud kujul kirjutada dokumendi karakteristika järgmiselt:

$$(m_{11} A) (m_{12} B) (m_{13} C) \dots (m_{in} K),$$

kus m_{11} , m_{12} , m_{13} , ..., m_{in} on mingi i-nda dokumendi karakteristikas sisalduvad märksõnad ja

¹ Черный А.И. Синтагматические отношения между дескрипторами. НТИ, 1969, сер. 2, № I, с. 7.

A, B, C, ..., K sümbolid märksõnade osatähtsuse väljendami-
seks.

Osatähtsuse näitajate kasutuselevõtmine tähendab informat-
sioonikeele täiendamist. Ühe märksõna m_1 asemel on võimalik kasu-
tada märksõnu m_1A , m_1B , m_1C , ..., m_1K , s.t. nii palju uue tähendusega märksõnu, kui palju osatähtsuse näitajaid kasutatakse antud süsteemis.

Leiutiste vormelite analüüsimisel selgus, et konstruktiivsete leiutiste puhul võivad esineda näiteks järgmised osatähtsuse näitajad:

A - leiutise objekt, s.t. peamine uurimise, konstrueerimise, väljatöötamise objekt (seade, masin, aparaat)

B - leiutise objekti koostisosa (konstruktiivne element), s.t. detailid, sõlmed, millest konstruktiivne leiutis koosneb

B_1 - konstruktiivse elemendi koostisosa

C - leiutise objekti materjal (aine)

C_1 - leiutise objekti koostisosa materjal (aine)

D - uudsuse tunnuseid omavad koostisosad

E - kasutamine - võimalik või eeldatav kasutusala, leiutise ülesanne.

Osatähtsuse näitajate hulk võib olla erinev ainevaldkonnast ja leiutise objektist (seade, meetod, aine).

Vaatleme eespool loetletud osatähtsuse näitajate kasutamist otsimistunnustikus, mis on koostatud leiutise kirjelduse nr.227997 vormeli põhjal.

Vormeli täielik tekst:

Seade käsiperfokaartide otsinguks, mis koosneb elektromagnetilisest vibraatorist, perfokaartide kastist ja kasti kinnitamise püstikust, erinev selle poolest, et perfokaartide otsingu kiirendamiseks ja töökindluse suurendamiseks on vibraator kujundatud kahest püstikust, elektromagnetitega, mille ankrud omavad eri-

nevaid võnkeamplituude, ja on seostatud telgede abil kasti vastaskülgedega.

Valime märksõnad ja lisame igale osatähtsuse näitaja:

| | |
|------------------------------|-------------------|
| käsiperfokaartide otsing | A |
| elektromagnetiline vibraator | B |
| kast | B |
| püstik | B, B ₁ |
| elektromagnet | B ₁ . |

Osatähtsuse näitajate kasutamine võimaldab otsingusüsteemi üles ehitada nii otsese kui inversioonimeetodiga. Viimase rakendamisel märksõnakaartide arv suureneb, kuna iga märksõna, mis esineb koos mingi osatähtsuse näitajaga, saab kartoteegis erineva kaardi.

Näiteks: elektromagnetiline vibraator A - leiutise objektina;

elektromagnetiline vibraator B - konstruktiivse koostisosana.

Osatähtsuse näitajate kasutamisel peetakse üldiselt peamiseks puuduseks seda, et indekseerijad võivad erinevalt, puhtsubjektiivselt määrata osatähtsuse näitajaid märksõnade juurde. Leiutiste vormelite analüüsimisel on subjektiivse teguri mõju minimaalne, kuna vormeli rangelt reeglipärane ülesehitus määrab üheselt märksõna loogilise funktsiooni.

Märksõnade osatähtsuse väljendamiseks võib otsimistunnustiku koostada ka märksõnade korteežidena, s.t. fikseeritud järjestusega märksõnade ahelana. Korteežis märksõna asetamine mingisse positsiooni määrab kindlaks tema funktsiooni analüüsitavas vormelis.

Otsimistunnustik esitatakse sel juhul tabelina või positsioonreferaadi kujul. Positsioonide arv võib olla kui tahes suur.

Leiutise kirjelduse otsimistunnustik võib olla esitatud näiteks järgmise tabelina.

| Positsioon, tema iseloomustus | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------------------------|-----------------|--------|------|---|----------------------|--|------------------|
| | leiutise objekt | | aine | protsess, milleks kasut. posits. 1,2,3 toodut | konstruks. elemendid | tehnol. protsessis kasutatavad komponendid | aine komponendid |
| | seade | meetod | | | | | |

Märksõnad

Kui dokumendi otsimistunnustikus puudub k-ndale positsioonile vastav märksõna, kirjutatakse tema asemel O.

Positsioonreferaadi koostamine võimaldab trükise sisu ühtset analüüsi ja otsimistunnustiku koostamist erinevate erialade spetsialistide poolt.

Märksõnade seose väljendamine dokumendi karakteristikas

Kui dokumendi iseloomustamiseks on valitud märksõnade hulk, siis võib otsimistunnustikus märksõnad rühmitada vastavalt analüüsitava dokumendi fragmentidele, mille sisu nad avavad. Spetsiaalse sümboli abil tähistatakse ära kokkukuuluvus.

Vaatleme märksõnade seoseid leiutise nr. 178 190 vormelis.

Vormeli täielik tekst:

Otsingusüsteem läbipaistvate visuaalkaartidega, erinev selle poolest, et süsteemi mahutavuse suurendamiseks, ilma kaartide mõõtmeid suurendamata, on läbipaistvad kaardid valmistatud formaatsete mikrofilmidena, millel on kujutatud originaalkaartide mikrokujutised.

Vormeli piiravas osas on olulisteks märksõnadeks: otsingusüsteem, läbipaistev, visuaalkaart.

Leiutise eritunnused: läbipaistev, visuaalkaart, formaatne mikrofilm, mikrokujutis.

Tähistame piirava osa hulka kuuluvaid märksõnu sümboliga S₁,

piirava osa märksõnu S_2 . Saame otsimistunnustiku kujul:

S_1 (otsingusüsteem, läbipaistev, visuaalkaart);

S_2 (läbipaistev, visuaalkaart, formaatne mikrofilm, mikrokuju-
jutis).

Võime esitada ka nii, et iga märksõna esineb otsimistunnus-
tikus üks kord. Siis saame ostimistunnustiku:

| | |
|---------------------|----------------|
| otsingusüsteem | - S_1 , |
| läbipaistev | - S_1, S_2 , |
| visuaalkaart | - S_1, S_2 , |
| formaatne mikrofilm | - S_2 , |
| mikrokujujutis | - S_2 . |

Kui tähistame seose näitajaid $S_1, S_2, i_3, \dots, i_k$, siis do-
kumendi otsimistunnustik võib omada üldise kaju

$i_1 (m_1, m_3, m_6)$

$i_2 (m_2, m_4)$

$i_k (m_1, m_2, \dots, m_n)$.

Seose näitajate kasutamisel koostatakse otsingusüsteem ta-
valiselt otsesel meetodil.

Loogiliste seoste väljendamine leiutise vormeli alusel koostatud otsimistunnustikus

Loogiliste seoste all mõeldakse antud juhul seoseid, mis täi-
davad nii osatähtsuse kui seose näitaja osa.

Leiutiste vormelite analüüs näitab, et iga leiutise vormelit
on võimalik vaadelda kui mingite elementide (objektide, esemete) ja
nende elementide vaheliste seoste kogumit.

Elementideks võib lugeda kõik kehad, ained, seadmed, masi-
nad, aparaadid, konstruktiivsed koostisosad, protsessid jne., s.t.
esemed, nende omadused, tegevuse.

Seoste e. relatsioonide all mõeldakse binaarseid seoseid elementide vahel.

Seosed, mis eksisteerivad elementide vahel, on väga mitmesugused, nii et neid tuleb vähendada, valida üldisemad ja tüüpilisemad. Liiga keerukas relatsioonide süsteem raskendaks otsimistunustiku koostamist. Relatsioonide normaliseerimist on uurinud G. Heinzmann ja Farradane. Viimane piirdub näiteks nelja relatsiooniga:

- A mõjutab B-d
- A = B
- A sisaldab B
- B lähtub A-st.

Allpool esitatud seoste süsteem on suure hulga leiutiste vormelite uurimise tulemus. Nad on sobivad ainult leiutiste vormelite analüüsimiseks.

Omavahel seotuna esinesid:

- E \longleftrightarrow E - ese esemega
- E \longleftrightarrow O - ese omadusega
- E \longleftrightarrow T - ese tegevusega
- O \longleftrightarrow O - omadus omadusega
- O \longleftrightarrow T - omadus tegevusega
- T \longleftrightarrow T - tegevus tegevusega.

Seostest olid leiutise vormeli jaoks tüüpilised järgmised:

- 0. A on identne B-ga R₀
- 1. A tuleneb B-st R₁
- 2. A määrab B otstarbe R₂

| | |
|---|----------|
| 3. A ja B | R_3 |
| 4. A või B | R_4 |
| 5. A mitte B | R_5 |
| 6. A koos B-ga või B koos A-ga | R_6 |
| 7. A sisaldab B või B sisaldab A | R_7 |
| 8. A omadusega B | R_8 |
| 9. A kaotab B | R_9 |
| 10. A mõjub B-le | R_{10} |
| 11. A enne B-d (ajaliselt) | R_{11} |
| 12. A pärast B-d (ajaliselt) | R_{12} |
| 13. A asetseb B ees (ruumiliselt) | R_{13} |
| 14. A asetseb B järel (ruumiliselt) | R_{14} |

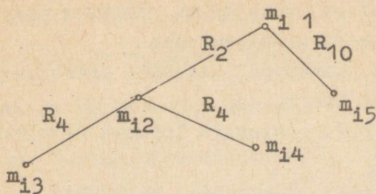
Kui igale seosele vastab kindel arv, siis võime otsimistunnustiku koostada märksõnade paaridest, kus märksõnad on seotud relatsiooniarvuga.

Näiteks kui mingit leiutist iseloomustab märksõnade hulk $m_{i1}, m_{i2}, \dots, m_{in}$ ja nende vahel esinevad seosed $1, 2, \dots, k$, siis otsimistunnustik võib omada kuju

$$\begin{array}{c}
 m_{i1}^{R_2} m_{i3} \\
 m_{i2}^{R_5} m_{i5} \\
 \dots \\
 m_{in-1}^{R_8} m_{in4}
 \end{array}$$

Me võime esitada märksõnadevahelisi seoseid ka graafiliselt, kus igale punktile vastab element leiutise vormelist ja neid punkte ühendavale sirglõigule vastab seos (relatsioon).

Näiteks kui vormelist valitud märksõnad on $m_{i1}, m_{i2}, m_{i3}, m_{i4}, m_{i5}$ ja seosed nende vahel vastavalt R_2, R_4, R_{10} , siis võib esitada otsimistunnustiku graafiliselt



Eespool toodud seoste hulka võib muidugi täiendada, osa seoseid võib osutada mittevajalikeks. Piisav ja vajalik seoste hulk selgub praktilises töös otsingusüsteemiga.

Märksõnadevaheliste loogiliste seoste väljendamine otsimistunnustikes võimaldab otsingusüsteemi koostada nii otsese kui inversioonimeetodi abil. Otseses süsteemis märgitakse igale informatsioonikandjale märksõnade paarid koos neid siduva relatsiooniga või kujutatakse seoseid graafiliselt.

Keerukam on otsingusüsteemi realiseerimine inversioonimeetodil.

5. INFORMATSIOONI OTSINGUSÜSTEEMI EFEKTIIVSUST MÕJUTAVAD TEGURID

Viimase kümne aasta jooksul on dokumentaalsete otsingusüsteemide efektiivsuse hindamiseks kasutatud mitmesuguseid meetodeid. Ulatuslikke uurimistöid on viidud läbi Leningradis V. Sokolovi juhendamisel, Inglismaal Krenfieldis U. Kleverdoni juhendamisel ja mujal.

Informatsiooni otsingusüsteemi kriitilist hindamist on vaja tavaliselt:

- uue süsteemi väljatöötamisel;
- rakendatud süsteemi hindamiseks;
- töötava süsteemi kvaliteedi pidevaks kontrolliks.

Põhieesmärgiks on süsteemi organiseerimine vastavalt informatsioonivajajate nõuetele.

Käesolevas kirjutises on tähelepanu pööratud peamiselt uue otsingusüsteemi väljatöötamise probleemidele.

Informatsioonivajajad hindavad süsteemi efektiivsust üldiselt järgmiste kriteeriumide järgi.

1. Süsteemi ulatus, s.t. süsteemis leiduva vajaliku kirjan-duse hulk.

2. Süsteemi võime leida küsimusele vastavaid, s.t. rele-vantseid dokumente.

3. Täpsus, s.t. võime ballasti (mitterelevantseid dokumente) mitte anda.

4. Otsinguaeg (ajavahemik küsimuse esitamise ja vastuse saa-mise vahel).

5. Vastuse vorm.

6. Informatsioonivajaja töö otsingul, s.t. jõupingutused, mida peab tegema küsija ise vastuse saamiseks.

Informatsioonivajaja poolt esitatud kriteeriumid on üks-teisega tihedalt seotud. Iga otsingusüsteem luuakse konkreet-se ülesande täitmiseks. Sellest sõltub ka oluliste kriteeriu-mide hulk. Mõningatel juhtudel on vajalik maksimaalne täielikkus (kriteerium 2), teistel juhtudel - maksimaalne täpsus (kritee-rium 3).

Otsingusüsteemi koostamisel, s.t. uue süsteemi väljatöö-tamisel tuleb arvestada neid tegureid, millest sõltuvad efek-tiivsuse kriteeriumid. Peale eespool toodute on kasutusel veel terve rida mitmesuguseid näitajaid otsingusüsteemi efektiivsu-se hindamiseks. Kõik need näitajad on saanud põhiliselt alljärgne-vas tabelis toodud suuruste alusel:

| | Relevantseid | | Mitterelevantseid | |
|-----------------------------------|--------------|-------|-------------------|-----------|
| | A | R | \bar{A} | \bar{R} |
| Otsingul leitud dokumendid | A | a | a | a + b |
| Otsingul mitteleitnud doku-mendid | \bar{A} | c | d | c + d |
| | | a + c | b + d | a+d+c+d |

Tabelis b väljendab ballasti, c - kadu

Otsingu täielikkus väljendub suhtena $\frac{a}{a+c}$

Otsingu täpsus väljendub suhtena $\frac{a}{a+b}$

Ballasti osa leitud dokumentidest $\frac{b}{a+b}$

Kao protsent väljendub suhtena $\frac{c}{a+c}$.

Uurimistöodes (vt. kirjanduse loetelus 8, 11, 12, 20) on jõutud järeldusele, et kõige enam võivad mõjutada otsingu tulemusi semantilised ja süntaktilised faktorid. Sellest võib teha järelduse, et kõige olulisema osa otsingusüsteemi koostamisel moodustab dokumentide analüüs, otsimistunnustiku koostamise meetodid.

Semantilistest faktoritest olulisemad on

- sünonüümsus,
- homonüümsus,
- paradigmaatilised seosed.

Märksõnade sünonüümsuse esinemine dokumentide otsimistunnustikes põhjustab informatsiooni kadu. Selle vältimiseks tuleb sisse viia viidete süsteem märksõnastikus või tesauruses.

Informatsiooni kadu põhjustab ka paradigmaatiliste seoste puudumine märksõnade vahel. Homonüümsuse esinemine annab otsingul ballasti.

Semantiliste tegurite mõju on võimalik reguleerida märksõnastiku või tesauruse koostamisel.

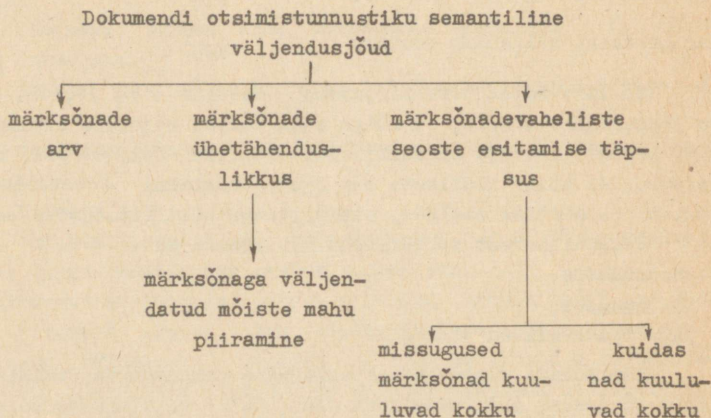
Süntaktilised faktorid mõjutavad otsingu täpsust. Süntagmaatiliste seoste kasutamine dokumentide otsimistunnustikes vähendab otsingul saadavat ballasti.

Mitmesuguste otsingusüsteemide uurimise tulemusena teeb A. Tšernõi oma artiklis "Deskriptoritevahelised süntagmaatilised seosed"¹ järelduse, et parimaks variandiks süntagmaatiliste seoste väljendamisel on loogiliste seoste süsteem, s.t. operaato-

¹ А. И. Черный. Синтагматические отношения между дескрипторами. НТИ, сер. 2, 1968, № 4, с.13.

rid, mis täidavad üheaegselt nii osatähtsuse kui seose näitaja osa.

Seega otsimistunnustiku kvaliteedi, tema semantilise väljendusjõu, vastavuse dokumendile määravad mitmesugused tegurid. Skeemaatiliselt võib seda kujutada järgmiselt:



Patendiinformatsiooni otsingusüsteemide koostamisel mõjutavad otsingu efektiivsust kõik eespool toodud tegurid.

Käesolevas töös peeti silmas kõigepealt patendikirjelduste otsimistunnustiku koostamise meetodeid.

Süntaktiliste faktorite reguleerimiseks on toodud meetodid lehekülgedel 17 kuni 29. Need meetodid on rakendatavad sõltumatult tehnikaharust. Semantiliste faktorite reguleerimiseks näib olevat kõige otstarbekam tesauruse koostamine. Tesauruse koostamise metoodikat antud töös ei käsitleta, kuna on silmas peetud, et käesolev otsingusüsteemi koostamise metoodika on sobiv nii lihtsa märksõnastiku, deskriptorsõnastiku kui ka keeruka ehitusega tesauruse korral.

Kui vastava eriala tesaurus puudub, on võimalik semantilisi tegureid reguleerida järgmiselt:

otsimistunnustiku koostamisel viia sisse täiendavalt üldisemad terminid;

informatsiooni otsingul viia läbi otsing ka antud märksõnadele alluvate terminite järgi;

otsida küsimuses esinevate märksõnade ja ka nende sünonüümide järgi.

Rahvusvahelist patendiklassifikatsiooni
kasutatavate riikide loetelu

- | | |
|--------------------|---------------------|
| 1. Albaania | 13. Prantsusmaa |
| 2. Austraalia | 14. Rootsi |
| 3. Belgia | 15. Rumeenia |
| 4. Bulgaaria | 16. Saksa DV |
| 5. Holland | 17. Saksa FV |
| 6. Iiri | 18. Soome |
| 7. Island | 19. Suurbritannia |
| 8. Itaalia | 20. Šveits |
| 9. Luksemburg | 21. Taani |
| 10. Norra | 22. Türgi |
| 11. Nõukogude Liit | 23. Tšehhoslovakkia |
| 12. Poola | 24. Ungari |

MÄRKSÖNA (UNITERM)

→ PERFOREERIMINE

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----|-----|-----|-----|-----|-----|-----|---|-----|-----|
| 10 | 31 | 2 | 13 | 14 | 25 | 6 | | 88 | 219 |
| 120 | 51 | 22 | 53 | 24 | 55 | 106 | | 218 | |
| 250 | 81 | 32 | 343 | 64 | 165 | | | 418 | |
| 420 | 91 | 62 | | 214 | 315 | | | | |
| | 111 | 102 | | | 525 | | | | |
| | 521 | 212 | | | | | | | |

Dokumentide numbrid,
mis seda märksõna sisaldavad

1. Блюмштейн Р. Возможности международного сотрудничества в области патентной информации и патентной документации. Доклад на заседании секции "Проблемы патентно-информационного поиска". М., 1969, (Научный симпозиум "Изобретательство и научно-технический прогресс").

2. Василевский А.Л. О некоторых особенностях языка патентной литературы. - В кн. "Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации", том I, М., 1967.

3. Вчерашний Р.П., Кравец Л.Г., Серх Г.Ф., Фурман Э.И. Патентная документация и патентная информация. М., ЦНИИПИ, 1969.

4. Добров Г.М., Колесников А.П. Вопросы тематического патентно-информационного поиска. - "Научно-техническая информация", 1966, № 12.

5. Кравец Л.Г. Зарубежный опыт механизации обработки и поиска патентной информации. - "Вопросы изобретательства", 1966, № 2.

6. Кравец Л.Г. Перспективы развития системы обработки и поиска патентной информации. Доклад на секции "Проблемы патентно-информационного поиска". М., 1969. (Научный симпозиум "Изобретательство и научно-технический прогресс").

7. Маркова Н.А., Шендеров В.З. Разработка пермутационного указателя для биллетеня "Изобретения, промышленные образцы, товарные знаки". - В кн. "Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации", том I, М., 1967.

8. Михайлов А.Ч., Черный А.И., Гиляревский Р.С. Основы информатики. М., "Наука", 1968.

9. Прогноз развития научно-технической информации до 1980 г. Основные направления в области методов и средств научно-технической информации. М., 1969, (Совет Экономической Взаимопомощи. Постоянная Комиссия по координации научных и технических исследований).

10. Указания по методике экспертизы заявок на изобретения (ЗЗ-2-67) .М., 1967.

11. Черный А.И. Синтагматические отношения между дескрипторами . - "Научно-техническая информация", 1968, Сер. 2, № 4.

12. Bader, K. -D. Probleme der Dokumentation von Patentschriften. - "Dokumentation/Information" II Kolloquium des Institutes für Dokumentation und Patentwesen an der TH Ilmenau vom 14. - 16. Nov. 1963. S. 61-71.

13. Hadlich, H. Gedanken über die Vorbereitung einer mechanisierten Auswahl von Patentliteratur mittels Deskriptoren. - "ZIID-Zeitschrift", 1966, 13, H.2, S.45-46.

14. Hoffmann, G. Ein allgemeines Informations-Recherche-System mit Hilfe Gerichter Graphen. - "Informatok", 1969, No 6, S. 19-20.

15. Kirchhöfer, D. Probleme der semantischen Information im Kommunikationsprozess - "ZIID-Zeitschrift", 15, 1968, H. 3, S. 109-112.

16. Meier, G. Linguistische, besonder semantische Probleme der Informationsverarbeitung. - "ZIID-Schriftenreihe", 12/1. S. 120-140. Berlin, 1966.

17. Rannap, E. Patendiinformatsioon. Tallinn, ENSV Informatsiooni Instituut, 1969.

18. Sharp, J.P. The SLIC Index. - "American Documentation", 1966, v. 17, No.1, p. 41-44.

19. Švankmajer, M. Der Grundtyp des Codes und die logische Struktur der Inhaltsregistrierung der Dokumente. - "ZIID-Schriftenreihe", 12/1, S. 160-167. Berlin, 1966.

20. Toman, J. Die verschiedene Methoden von Informationsverarbeitung. - "Nachrichten für Dokumentation", 1965, H. 3, S. 117-124.

Sisukord

| | |
|--|----|
| Sissejuhatus | 3 |
| 1. Leiutise kirjeldus informatsiooniallikana | 5 |
| 2. Otsingusüsteemide koostamine | 8 |
| 3. Leiutiste kirjelduste analüüs | 12 |
| 4. Leiutise vormeli analüüs | 15 |
| 4.1. Märksõnade (unitermide) valik leiutise vormelist | 15 |
| 4.2. Süntagmaatilised seosed leiutise vormelis | 17 |
| 5. Informatsiooni otsingusüsteemi efektiivsust mõjutavad tegurid | 29 |
| 6. Lisad | 34 |
| 7. Kirjandus | 37 |

Эви Р а н н а п
О методах анализа описаний
изобретений

На восточном языке
Издательство "Валгус"
Таллин, Пярнуское шоссе, 10

Toimetajad M.Moora ja P.Sarevet

Trükkimisele antud 21.08.70. Faber 60x84/16.
Trükiplaanid 2,5 + 1 kleebis. Tingtrükiplaanid
2,44. Arvestusplaanid 2,12. MB - 05954
Trükiarv 550. Tellimise nr. 1458-1549.
Eksperimentaalkombinaadi "Bit" rotaprint,
Tallinn, Pikk 68.

Hind 21 kop.

Hind 21 kop.