

TESAURUSTE LOOMISE TEOORIAST JA PRAKTIKAST

J. Valge

TRÜ Kriminoloogia Laboratoorium

O. Õiguse kasvav tähtsus nii sotsiaal-majanduslikus kui poliitilises elus, suurenev nõudmine operatiivse ja ammendava juriidilise informatsiooni järele nõuavad õigusliku reguleerimise kaasajastamist ning õigusealase informatsiooniteenistuse parandamist (A. P. Шляхов, С. С. Москвин, 1973).

Viimase ülesandega on otseselt seotud TRÜ Kriminoloogia Laboratooriumi õigusliku informatsiooni probleemgrupis teostatav töö. See vahest olekski põhjuseks, mis ajendas vaatlema mõningaid tesauruste — seni viimane ja moodsaim väljakujunenud informatsioonikeel — loomisega seotud probleeme käesoleva kogumiku veergudel.

O. 1. Tesauruste ulatuslik koostamine¹ algas viiekümnendate aastate lõpul ja on üha elavnevas tempos kestnud tänapäevani. Loomulikult on tänaste ja aastaid tagasi loodud tesauruste vahel palju erinevusi: mitmed probleemid, mis siis püstitati, on nüüd lahendatud; samal ajal on tekkinud palju uusi vastust nõudvaid küsimusi.

On muutunud ka tesauruste loomise teooria ja praktika vaherkord. Kui viiekümnendatel aastatel käis teooria praktikast kaugel ees, viimane aga taotles talle teoreetikute poolt esitatud nõuete täitmist, siis tänapäeval on mõlemad märgatavalt iseseisvunud. Seoses sellega võib üha rohkem kuulda hoiatavaid hääli üldisest edasiviivast teooriast loobumise ja kitsalt pragmaatilistel kaalutlustel rajanevate erialatesauruste loomise aadressil.

On selge, et informatsiooni otsimise süsteemi osana on tesaurused suure praktilise tähtsusega. Sama selge aga on ka see, et ainult eesrindlikest ja arendatud teooriatest lähtumine võimaldab tesaurustel rahuldada neile esitatavaid üha kasvavaid nõudmisi.

¹ Siinkohal eeldatakse, et artikli lugejale on tesaurusega seotud põhimõisted tuttavad. Vastasel korral on soovitatav lugeda monograafia «Основы информатики» tesaurusi käsitlevat peatükki.

Järgnevalt püüamegi vaadelda pärast 1961. aastat ilmunud tesaurustele iseloomulikke jooni. Seejuures pole eesmärgiks mitte üksikute tesauruste täpne kirjeldamine, vaid üldiste tendentside näitamine.

Ülevaate koostamiseks on kasutatud väga mitmesuguseid allikmaterjale, neist põhilistena monograafiat «Основы информатики» ja referatiivajakirja «Информатика».

Üksikprobleemide selgituseks kasutatud täiendavatele allikatele viidatakse edaspidi.

Tinglikult võib tesaurusi jaotada kahte rühma: 1) keele üldist semantilist struktuuri iseloomustavad ja 2) teatud dokumentide massiivis informatsiooni otsimiseks määratud tesaurused. Esimeste ülesandeks on rohkete sünonüümide esiletoomine, väljenduse mitmekesistamisele kaasaaitamine; teised, vastupidi, peavad kindlustama sünonüümia kaotamise, sisu võimalikult täpse ja ühemõttelise edasiandmise.

0.2. Käesolevas töös vaatleme ülevaاتlikult järgmisi kitsamaid probleeme (praegune artikkel on TRÜ Kriminoloogia Laboratooriumi juriidilise informatsiooni probleemgrupis peetud samatemaalise ettekande arendus):

- 1) tesauruste liigitamise alused, tesauruste ülesanded;
- 2) tesauruse struktuur;
- 3) tesauruse maht, selle kujunemine;
- 4) tesauruste loomine: kus, kelle poolt, mis aladel ja kuidas toimub tänapäeval tesauruste koostamine;
- 5) tesauruste loomise koordineerimine;
- 6) mõningaid aktuaalsemaid probleeme seoses tesauruste koostamisega;
- 7) juriidilise tesauruse koostamisest TRÜ Kriminoloogia Laboratooriumis.

1. Koos tesauruste arvu suurenemisega kasvab vajadus neid liigitada. Tunnuseid, mille põhjal seda tehakse, on mitmeid. Esitame siin ühe Saksa DV-s (kuna just seal on tesauruste teooria põhjalikult läbi töötatud) avaldatud liigitussüsteemi.

- 1.1. Viidatud süsteemis võime rääkida
 - a) vormilt (alfabeetiline või süstemaatiline),
 - b) sisult (spetsiaalne või üldine),
 - c) deskriptorite tüübilt (mõisted, arvulised suurused, graafilised sümbolid),
 - d) seostelt erinevatest tesaurustest.

1.2. Mainigem, et tesaurusi liigitatakse ka selle järgi, kuidas nad on koostatud. Koostamisviisist sõltuvalt võime rääkida

- a) lingvistilistest tesaurustest (koostatakse tekstide sisulise analüüsi põhjal);

b) normatiivsetest tesaurustest (seosed määratakse aprior-selt);

c) statistilistest tesaurustest (seosed määratakse tekstide sta-tistilise analüüsi põhjal).

Esimest kahte tüüpi võib nimetada eksperdi hinnangu mee-todil (метод экспертной оценки) loodud tesaurusteks; statistili-se tesaurused saadakse automaatse analüüsi abil.

Esimeste puuduseks on suur ajakulu koostamisel ning sellest tulenev ebaoperatiivsus, samuti tulemuste subjektiivsus.

Teiste puhul aga tuleb silmas pidada, et arvuti annab vaid interpreteerimiseks materjali, suutmata seda ise mõtestada.

1.3. Olgu lisatud, et tesaurust ei tule vaadelda sugugi ainult teatud informatsioonisüsteemi osana (nagu juba eespool viidatud).

Kuigi tesauruste põhiliseks ülesandeks on vastava ala doku-mentide indekseerimisele ja otsimisele kaasaitamine, on ka teisi kasutusvõimalusi.

On tesaurusi, mis pole loodud ühegi konkreetse informatsiooni-süsteemi jaoks, vaid terminoloogia unifikseerimiseks, abivahendiks nii referaatide kui ka esmaste dokumentide eneste (näiteks sea-dusandlike aktide) koostamisel.

Lisagem, et kui tesauruste loomise algaastail nähti ainult nende rakenduslikke eesmärgi, siis tänapäeval mõistetakse, et tesaurus on ikkagi vastava ala keele semantika kirjeldamise üks viis. Siit tuleneb tesauruste lingvistiline tähtsus.

2. Rääkides tesauruse struktuurist, tuleb lähtuda mitmete nimekate spetsialistide poolt väljendatud seisukohast, et tesau-ruse struktuur, liigendus on palju tähtsam kui üksikute deskrip-torite ja nende seoste vastavus tegelikkusele. Järgnevalt vaatle-megi, milline on tesauruste struktuur ja milliseid nõudeid sellele esitatakse.

2.1. Klassikalise tesauruse põhilisteks osadeks on deskripto-rite alfabeetiline loend ja teatud süsteemis näidatud seosed nende deskriptorite vahel.

Mitmetes tesaurustes esitatakse koos lubatud deskriptoritega ka keelatud deskriptorid, s. o. sõnad, mis võivad küll esineda doku-mentides, kuid mida ei kasutata indekseerimisel ega informatsiooni otsimisel.

Praktikas on ilmselt kasulik lülitada tesaurusesse ka mitme-suguste asutuste, geograafiliste punktide jms. nimetused. Mõnedel juhtudel on toponüümiline materjal tesauruse eri lisana välja antud. TRU Kriminoloogia Laboratooriumil on kavatsus koostada eraldi asutuste, organisatsioonide nimetuste tesaurus.

Tesauruses esitatavatest seostest on tähtsaimad hierarhilised, s. o. soo-liigi ja osa-terviku seosed. Viimasel ajal räägitakse sageli assotsiatiivse seoste tähtsusest. Sellele tendentsile on kind-lasti soodustavalt mõjunud arvute rakendamine tesauruste loo-

misel, sest tekstide automaatsel analüüsil selguvad eelkõige assotsiatiivsed seosed.

2.2. Tänapäeva tesaurusi iseloomustab mitmesuguste nende kasutamist hõlbustavate lisade rohkus. Võib nähtavasti väita, et just lisade olemasolu või puudumise põhjal tehakse vahet, kas on tegemist tesauruse või deskriptorsõnastikuga (neid mõisteid kiputakse samastama). Deskriptorsõnastik on teatud kindla informatsioonisüsteemi osa, milles esitatakse vaid deskriptorite loend ja seosed nende vahel; tesauruse all aga mõistame enamasti eraldi väljaantud, hoopis laiemal rakendamispiirkonnaga käsiraamatut.

Mainitud lisades antakse põhjalik instruksioon tesauruse kasutamiseks, selgitatakse seda näidetega. Esitatakse mitmesuguseid registreid ning sageli tutvustatakse ka konkreetse tesauruse loomise metoodikat ja koostamisega seoses olevaid organisatsioonilisi küsimusi.

2.3. Üheks olulisemaks tesauruste struktuurile esitatavaks nõudeks on paindlikkus. Mitmete tesauruste lisades ja tutvustavates-retsenseerivates artiklites peetakse tesauruse täiendamise ja parandamise võimalust üheks tema hindamise põhiliseks kriteeriumiks. On ilmne, et suvalise tesauruse areng kestab vähemalt nii kaua, kuni temas käsitletav ala areneb ja rikastub uue terminoloogia ning uute sisuliste aspektidega.

2.4. Nii nagu tänapäeva teaduslikku ja tehnilist tegevust ning sotsiaalseid juhtimis- ja organiseerimisprotsesse üldse, nii iseloomustab ka informatsiooni otsimist ja töötlemist raalide laialdane kasutamine ülalnimetatud tegevusalade tarbeks. Siit tuleneb veel üks tesauruste struktuurile esitatav nõue: rakendatavus automaatsetes informatsiooni otsimise ja töötlemise süsteemides.

2.5. Tänapäeval areneb tesauruste loomine põhiliselt kahe suunas: ühelt poolt luuakse suuri, polütemaatilisi tesaurusi ja tesauruste süsteeme, teiselt poolt aga kasutatakse üha rohkem kitsalt spetsialiseeritud erialatesaurusi.

Üldiseks nõudeks on mahukate, laia temaatikat haaravate tesauruste suur üldistusjõud, terminoloogia fikseeritus ja universaalne struktuur.

Spetsialiseeritud tesaurustes lubatakse mõnevõrra vabamat ümberkäimist terminitega, kohati tuuakse sisse isegi dokumentides esinevaid žargonisme. Struktuur, mis peab kattuma vastavat ala hõlmava üldise tesauruse struktuuriga, on detailiseeritum. Soovitatakse kasutada üldiste tesauruste vastavalt kohandatud sõnartikleid.

3. Tesauruste mahust rääkides mainigem kõigepealt, et see pole tesauruse hindamise kriteeriumiks. Iseenesest pole tähtis, kui suur või väike on tesaurus. Oluline on, et ta vastaks esitatavatele nõuetele.

On olemas väga mitmesuguse, temaatikast ja teistest informatsioonisüsteemi iseärasustest oleneva mahuga tesaurusi.

Tesauruste kirjeldustes esitatakse tavaliselt arvulised andmed kahe suuruse kohta:

- a) mitu deskriptorit sisaldub antud tesauruses;
- b) mitme dokumendi läbitöötamise tulemusena see tesaurus on valminud.

Mõnikord viidatakse ka seoste arvule.

Järgnevalt mõned näited tesauruste arvuliste karakteristikute kohta.

3.1. Suurimate tesauruste hulka kuuluvad USA-s avaldatud valitsuse dokumentide massiivi tesaurus (32 842 deskriptorit) ning «Tehniliste ja teaduslike terminite tesaurus» (23 364 deskriptorit), Euratomi tesaurus (16 473 deskriptorit), Norras koostatud polütehniliste terminite tesaurus (18 000 deskriptorit).

On aga ka palju kordi väiksemaid tesaurusi. Slovaki TA Orgaanilise Keemia Instituudi tesaurus sisaldab 540, NSV Liidus koostatud deskriptorsõnastik torujuhtmetranspordi alal 562 deskriptorit jne.

Esitatud arvulistesse väärtustesse tuleb muidugi suhtuda ettevaatlikult, sest nagu R. Nigol on maininud (R. Nigol, 1972), leidub informaatika-alases terminoloogias suuri lahknemisi. Selle tunnistuseks on asjaolu, et ilmub kümnete kaupa artikleid, mille sisuks on selliste põhimõistete nagu «indekseerimine», «deskriptor», «tesaurus» jne. defineerimine. Ja on muidugi selge, et need definitsioonid ei kattu.

3.2. Näitena tesauruse mahu muutumise kohta vastavalt analüüsitud dokumentide arvule võib tuua USA-s koostatud «Isotoopidealase informatsiooni tsentri võtmesõnade tesauruse», kus võtmesõnade arv kujunes järgmiste etappide kaupa: 1000 dokumenti — 39 võtmesõna, 3500 dokumenti — 165 võtmesõna ja lõpuks 7000 dokumenti — 200 võtmesõna.

Muutus ka Belgia firma «Solvay» plastmassikeemia-alane tesaurus. Algul koosnes see 2000 dokumendi põhjal loodud tesaurus 4000 terminist; 4 aasta vältel vaadati läbi veel umbes 50 000 dokumenti, mille järel sõnastik kasvas 8500 terminini. Seejärel sõnastiku kasv praktiliselt lakkas.

4. Klassifikatsioonide järel on tesaurused teiseks informatsioonikeelte põlvkonnaks. Sarnaselt enamiku avastuste ja leiutistega on tesaurustegi tekke otseseks stimulaatoriks vajadus nende järele.

4.1. Tesaurusi hakati looma eelkõige seal, kus teadusliku ja tehnilise informatsiooni maht saavutas kõige kõrgema taseme ning kus tekkis vajadus seda informatsiooni operatiivselt kasutada.

Kuigi lähtekirjanduse ilmselt mittetäielikud andmed ei anna täiesti realistlikku pilti eri maade panusest olemasolevate ja valmimisjärgus tesauruste fondi, võib väita, et nende koostamise raskuspunktid on Ameerikas (eelkõige USA, aga ka Kanada), Euroopas (Inglismaa, Prantsusmaa, Belgia, Saksa FV; samuti ka

sotsialismimaad eesotsas NSV Liidu ja Saksa DV-ga) ja Jaapanis.

Oluline koht tesauruste loomisel on ka rahvusvahelistel organisatsioonidel ja riikidevahelise koostöö teistel vormidel.

Tesauruste otsesteks koostajateks võivad olla nende järelevajadust tundvad firmad (eriti kapitalistlikes riikides — näiteks firma «Solvay» Belgias), riiklikud ja muud organisatsioonid (NASA), spetsiaalsed informatsiooniinstituudid ja -komiteed (ВИНИТИ).

Sageli võtavad tesauruste loomisest kas siis lepingute alusel või teadusliku uurimise otstarbel osa ülikoolid.

4.2. Väga mitmesugused on ka alad, mille kohta tesaurusi koostatakse. Siingi tingib tesauruse loomise eelkõige vajadus tema järele. Kõige enam tesaurusi on koostatud nendel teaduse ja tehnika aladel, mis arenevad tänapäeval kõige kiiremini. Nime tagem selliseid alasid nagu keemia ja keemiatööstus, kosmonautika ja elektrotehnika, tuumafüüsika ja meditsiin, samuti põllumajandus ja transport. Mõõda ei saa minna tesaurustest, mis käsitlevad haridust, ehitust, tööstuse juhtimist ja veel mitmeid teisi valdkondi.

On välja kujunenud riigid, millele kuulub prioriteet ühe või teise teadusharu tesauruste koostamisel.

Nagu mainitud aladest näha, luuakse enamik tesaurusi looduse ja täppisteaduste ning mitmesuguste tootmisharude kohta. Humanitaarteaduslike dokumentide indekseerimiseks ja otsimiseks on loodud suhteliselt vähe tesaurusi. Ilmselt raskendab ühiskonnateadusi käsitlevate tesauruste loomist terminoloogia suhteliselt vabam liikumine, seoste sagedane mitmetähenduslikkus.

4.3. Mis puutub tesauruse koostamise viisi, siis võib üpris kindlalt väita, et põhiliseks tesauruste loomise meetodiks on teatud ala dokumentide käsitsi läbitöötamine ja selle põhjal sobivate märksõnade ning nendevaheliste seoste leidmine. Muidugi võidakse mõnedes tehnilise töö lõikudes seejuures kasutada (ja kasutataksegi) raali. Sellel meetodil on koostatud kõik suuremad ja üldist tunnustust leidnud (autorile teada olevad) tesaurused.

Siiski on juba viiekümnendate aastate keskelt arenenud ka põhimõtteliselt teine suund — tesauruste automaatne koostamine. Selle aluseks on nii teistsiti tekstides esinevate sõnade statistiliste karakteristikute (üksikute sõnade esinemissagedus, sõnade koosesinemise sagedus jne.) arvestamine (vt. lähemalt J. Valge, 1972).

Viimasel ajal on jõutud tesauruste koostamise distributiiv-statistilise meetodi loomise ja rakendamiseni (B. A. Москович, 1971). Meetod põhineb eeldusel, et kui kaks sõna esinevad tekstis koos sagedamini kui teineteisest sõltumatute sündmuste puhul tõenäoline oleks, siis on nende sõnade vahel mingi mõtteline seos. Muu-

tes selle tekstilõigu pikkust, mille piires me koosinemist vaatleme, võime saada erinevat tüüpi seoseid. On muidugi selge, et põhilised seosed, mis sellise meetodiga leitakse, on assotsiatiivset laadi.

Lisagem veel seda, et distributiiv-statistiline meetod eeldab sõnadevaheliste seoste tugevuse mõõtmist, mille põhjal saab eraldada olulised seosed vähefäätsatest.

Parimaid tulemusi on nähtavasti võimalik saavutada mõlemaid meetodeid kombineerides: tesaurus koostatakse käsitsemetodil, distributiiv-statistiline analüüs oleks seejärel täiendamising kontrollivahendiks.

5. Informatsioonihulga järsk suurenemine kõikidel teaduse ja tehnika aladel ning sellest tulenev tesauruste mahu ning maksimuse kasv on päevakorda tõstnud tesauruste-alase töö koordineerimise. Üksikud organisatsioonid pole lihtsalt võimelised looma tesaurust, mis valmimise hetkel poleks juba moraalselt vananenud (seda eriti kiiresti arenevates teadusharudes).

5.1. Koostöö sellel alal meil, sotsialistlikes riikides, peaks olema põhimõtteliselt lihtne: juhivad ju sellealast tööd mitmesugused informatsiooniinstituudid ja -komiteed.

Eriti põhjalikult tegeldakse tesauruste-alase töö koordineerimisega Saksa DV-s, kus kõikidest arendatavaist suundadest on ülevaade ja kus neid tsentraalselt juhitakse.

Kõrvalmärkusena olgu aga lisatud, et hoolimata koostöö põhimõttelisest lihtsusest põrkavad sellelaadsed püüdlused sageli takistustele.

Koostöö näiteid on küllalt ka kapitalistlike riikide kohta. Näiteks USA-s võttis haridusalase tesauruse loomisest osa ligi paarikümmend asjast huvitatud organisatsiooni.

Endastmõistetav on ka eri riikide koostöö rahvusvaheliste organisatsioonide raames, mille heaks näiteks on Euratomi tesaurus.

5.2. Lisaks otsesele osavõtule tesauruse loomisest on mitmeid teisigi võimalusi selle täiustamisele kaasaaitamiseks.

Koostööd vastava ala spetsialistidega valmimisjärgus oleva tesauruse täiendamiseks ja parandamiseks rakendatakse ka TRU Kriminoloogia Laboratooriumis.

Väga levinud on nii valminud tesauruste kui nende katsekemplaride retsenseerimine erialases kirjanduses.

Tesaurust täiendatakse sageli tema kasutajatelt saadud hinnanguate põhjal. Näiteks võib tuua USA informatsioonikeskustes ja raamatukogudes teostatud ankeedi NASA tesauruse ning «Tehniliste ja teaduslike terminite tesauruse» kohta. Küsimused puudutasid tesauruse struktuuri, kasutamise sobivust j.m. probleeme. Vastused üldistati ja neid kasutatakse edaspidi vastavate tesauruste täiustamiseks.

5.3. Vaadeldes koostöö võimalusi alade kaupa, on selge, et lihtsam on see täpse, fikseeritud terminoloogiaga teaduse ja tehnika harudes. Isegi juhul, kui terminid vormilt ei kattu, on võimalik üldise süsteemi loomine sisu alusel.

Teine lugu on aga mitmesuguste ühiskonnateadustega (sealhulgas ka õigusteadusega), kus kujult sarnaste terminite taga võib peituda küllaltki erinev sisu.

Seoses sellega võib märkida Saksa DV-s esitatud seisukohta poliitilise ökonomia tesauruse kohta: nimelt soovitatakse see luua eraldi nii kapitalismi kui sotsialismi majandusteaduse jaoks.

5.4. Viimasel ajal võib rääkida veel ühest võimalusest tesaurustealaseks koostööks. Nimelt on tesauruste tase saavutanud sellise kõrguse, et tesaurus ise (mitte ainult selle loomine) on muutunud teadusliku uurimise objektiks (mitmed tööd selle kohta on ilmunud SDV-s ja Inglismaal). Koostöö sellel tasemel aitaks kindlasti kaasa tulevaste tesauruste otstarbekamale koostamisele.

Küsimus on seda aktuaalsem, et Poolas näidati meditsiinilise tesauruse kasutamist uurides, kuidas tesaurus mõjutab teda rakendava asutuse või organisatsiooni struktuuri.

6. Tänapäeval on tesauruste loomise teoorias ja praktikas mitmeid kiiret lahendamist nõudvaid probleeme. Nimetagem lühidalt neist tähtsamaid.

Teaduse ja tootmise internatsionaalse iseloomu tõttu on väga aktuaalne terminoloogia unifikseerimise ja üks-, kaks- (ning mitme-)keelsete tesauruste loomise üldiste printsiipide väljatöötamine.

Rahvusvahelise koostöö aluseks saavad ilmselt UNESCO poolt soovitatud üks- ja kakskeelsete tesauruste koostamise põhiprintsiibid. Need on esitatud väga põhjalikena, eesmärgiga võimaldada erinevate rahvuslike informatsiooniteenistuste koostööd.

Tuuakse tesauruse loomisega seoses olevate terminite määratlused, samuti seoste liigid, mis peavad tesauruses sisalduma. Nõutakse süsteemi muutmise võimalust: aegunud terminite väljajätmist ja uute lisamist.

Kakskeelsete tesauruste koostamine on ilmselt veel niivõrd uus ja kujunemata ala, et UNESCO põhiprintsiibidki on formuleeritud esialgseina. Igal juhul soovitatakse kakskeelse tesauruse koostamist alustada baaskeelse tesauruse korrigeerimisest.

Lisatagu, et kakskeelse tesauruse (eesti-vene) koostamise võimalusi on vaadeldud ka TRÜ Kriminoloogia Laboratooriumis. Tuleb ju meie informatsioonisüsteemil vastata mõlemas keeles esitatud tellimustele.

7. Lõpetuseks kirjeldatagu mõne sõnaga TRÜ Kriminoloogia Laboratooriumis loodavat tesaurust (vt. lähemalt И. Г. Кулль, И. Я. Сильдмяэ и др., 1973).

7.1. Tegemist on lingvistilise, juriidiliste tekstide käsitsi läbitöötamise tulemusena koostatava tesaurusega. Lisaks otseselt

juriidilistele terminitele sisaldab see ka nende alade termineid, kus õiguslik reguleerimine toimub.

Tesaurusse korrastamisel ja lisaregistrite koostamisel kasutatakse raali abi.

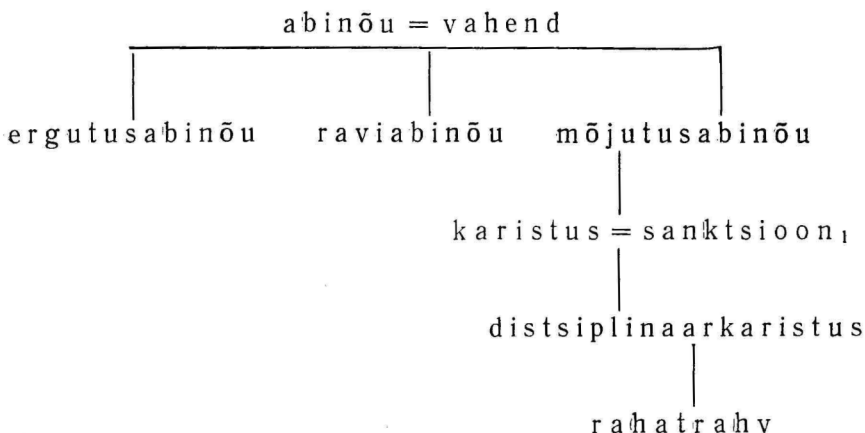
Tesaurus on mõeldud automatiseeritud informatsioonisüsteemi osana, samuti aga ka tunduvalt laiemal otstarbel kasutamiseks.

Lõplikul kujul peaks tesaurus sisaldama umbes 15000 märksõna koos seostega.

7.2. Tesaurus esitatakse järgmised seosed mõistete vahel.

7.2.1. Sünonüümia. On küllalt juhtusid, kus ühe või teise mõiste väljendamiseks kasutatakse juriidilistes tekstides mitut eri sõna. Õigete dokumentide efektiivseks otsimiseks võetakse kõik need sõnad tesaurusesse, viidates seejuures põhisõnale (ühele neist). Põhisõna valikul arvestatakse vastava termini levikut, samuti perspektiivi. Nii näiteks on «iseteeninduskaupluse» sünonüümiks (ja põhisõnaks) võetud «selvekauplus».

7.2.2. Sugu-liik. Selle seose esiletoomiseks kasutatakse moodust, kus iga mõiste juures esitatakse vaid temast vahetult astme võrra kõrgemal ja madalamal olevad mõisted. Kogu süsteemi raames on selle põhjal muidugi võimalik ulatuslike hierarhiate loomine. Esitatagu siin üks neist (on muidugi võimalikud mitmed paralleelsed hierarhiad):



7.2.3. Osa-tervik. Sama süsteemi järgi seatakse (vahetu eelnevus ja järgnevus) mõisted ka siin. Näiteks «advokaat» ∈ «advokaatide kolleegium».

7.2.4. Eitus. Esitatud märksõnale tuuakse juurde tema vastandmõiste. Näiteks: «puhkepäev» ≠ «tööpäev».

7.2.5. Funktsionaalne seos. Funktsionaalselt seotakse mõisted, millest üks tähistab teise funktsiooni või tulemust ja vastupidi. Näiteks «kurjategija» → «kuritegu»; «loeng» ← «lektor».

7.2.6. Assotsiatiivne seos. Assotsiatiivselt seostatakse mõistet, mida pole võimalik siduda juba mainitud seoste abil, kuid mille sidumine informatsiooni otsimise seisukohalt võiks olla vajalik. Näiteks «nuumamine» ~ «loomakasvatuse», «naistööjõud» ~ «naisisik».

7.2.7. Oluline juriidiline seos. Nii seotakse omavahel assotsiatiivses seoses olevaid mõisteid, millel on eriline juriidiline, informatsiooni otsimise seisukohast vajalik tagapõhi. Näiteks «töötasu» \approx «stipendium», \approx «pension».

Viimased pole mingil määral töötasu liigid, kuid mitmes juriidiliselt olulises aspektis (näit. maksude arvestamisel) on neil sarnaseid jooni.

8. Loodetavasti andis eelnev ülevaade pildi tesauruste loomisest tänapäeval ja samuti valgustas TRU Kriminoloogia Laboratooriumis tehtavat tööd.

KIRJANDUS

- Куль, И. Г., Сильдмяэ, И. Я., Хелемяэ, А. К., Ыйм, Х. Я. 1973. О разработке тезауруса юридических терминов для информационно-поисковой системы. — Правовая кибернетика, М., стр. 54—62.
- Михайлов, А. И., Черный, А. И., Гиляревский, Р. С. 1968. Основы информатики, М.
- Москович, В. А. 1971. Дистрибутивно-статистический метод построения тезаурусов: Современное состояние и перспективы, М.
- Шляхов, А. Р., Москвин, С. С. 1973. Актуальные вопросы совершенствования учета и поиска нормативных материалов. — Правовая кибернетика, М., стр. 22—34.
- Nigol, R. 1972. Juriidilise informatsiooni töötlemise automatiseeritud süsteemide loomisest lääneriikides. — Käsikiri TRU Kriminoloogia Laboratooriumis.
- Valge, J. 1972. Sõnadevaheliste semantiliste seoste kindlaksmääramine keelestatistika meetoditega. — Käsikiri TRU Kriminoloogia Laboratooriumis.
- Valge, J. 1973. Mõningaid probleeme tänapäeva teauruste loomisest teoorias ja praktikas. — Käsikiri TRU Kriminoloogia Laboratooriumis.

О ТЕОРИИ И ПРАКТИКЕ СОЗДАНИЯ ТЕЗАУРУСОВ

Ю. Валге

Резюме

Одновременно с быстрым увеличением количества научной и технической информации растет и значение тезаурусов.

При создании тезаурусов имеются некоторые общие тенденции, которые следует иметь в виду при создании каждого тезауруса.

Для иллюстрации проблем, связанных с созданием юридического тезауруса в Лаборатории криминологии ТГУ, в настоящей работе анализируются следующие вопросы:

- 1) основы разделения тезаурусов, задачи тезаурусов;
- 2) структура тезауруса;
- 3) объем тезауруса и образование его;
- 4) создание тезаурусов (где, кем и в какой области происходит создание тезаурусов);
- 5) координация создания тезаурусов;
- 6) юридический тезаурус, создаваемый в Лаборатории криминологии ТГУ.