

RELEVANTIŠKUMO SĄVOKOS RAIDA
REGINA JUREVICIOTĖ

Informatikoje svarbi ir daug diskusijų sukėlusi yra relevantiškumo* sąvoka. Žymus informatikas V. Kuperis rašė: „Atrodo, bendrai sutariama..., kad informacijos paieškos procesas niekada nebus iki galo suprastas, prieš tai nesupratus tos neaiškios sąvokos, kuri vadinama relevantiškumu. Relevantiškumas yra vieno iš pagrindinių, jeigu tik ne pagrindinė, informacijos paieškos teorijos sąvoka“¹.

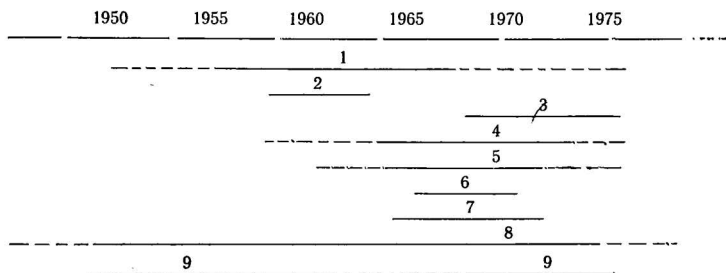
Relevantiškumas — tai dokumento turinio atitikimas informacinę užklausa. Užklausa atitinkantys dokumentai vadinami relevantiškais. Išduoti rele-

vantiškus dokumentus — svarbus informacijos paieškos sistemų (IPS) uždavinys. Todėl relevantiškumo problemų tyrimas ir išsprendimas turi didelę reikšmę informacijos paieškos procesų tobulinimui. Apie įvairius jo aspektus ir problemas yra pasirodžiusi gausi literatūra. Remiantis ja, straipsnyje siekiama išanalizuoti relevantiškumo sąvokos raidą ir jos tyrimo kryptis iki aštuntojo dešimtmečio.

Svarbiausius relevantiškumo sąvokos raidos etapus (iki 1970 metų) ištyrė ir apibūdino T. Saracevičius² (žr. schemą Nr. I).

Schema Nr. I.

PAGRINDINIAI RELEVANTIŠKUMO SĄVOKOS RAIDOS ETAPAI



1. Relevantiškumo problemos iškilimas. 2. Pirmosios diskusijos. 3. Teorių kūrimas. 4. Vertinimo matų, relevantiškumą laikančių kriterijumi, kūrimas. 5. Relevantiškumo panaudojimas ir tyrimas IPS eksperimentuose. 6. Antrasis diskusijų etapas. 7. Apibrėžimai ir hipotezės. 8. Specializuoti relevantiškumo eksperimentai. 9. Relevantiškų dokumentų distribucijos tyrimai.

* Rengiant straipsnį spaudai, terminas „relevantiškumas“ pakeistas terminu „relevantumas“.

¹ Cooper W. S. A. definition of relevance in information retrieval. — Information Storage and Retrieval, 1971, v. 7, No 1, p. 19.

² Saracevic T. The concept of relevance in information science: A historical review. — In the book: Introduction to Information Science. New-York—London, 1970, p. 111—151.

Jis pažymėjo, kad iki to laiko, kai apie relevantiškumą pradėta kalbėti informatikoje, jo sąvoka visada egzistavo moksle, būdama vienu iš pagrindinių kriterijų: tyrinėjant kurį nors objektą, nuolat ieškoma jam pritaikomų, t. y. relevantiškų, rezultatų kitose, šiam objektui giminingose ir svetimose srityse. Relevantiškumo sąvoka buvo pastebėta ir aptarta jau senovės filosofų. Tačiau relevantiškumo suvokimas informacijos paieškos teorijoje — ne filosofinės kilmės. Jis atsirado, bibliotekų ir IPS eksploatavimo praktikoje pradėjus naudoti ESM. Nuo to laiko supratimas apie relevantiškumo esmę praėjo daugelį mokslo vystymuisi būdingų etapų: buvo suvoktas reiškinio egzistavimas, iškeltos jo prigimties prielaidos, pastebėti šių prielaidų trūkumai, pasiūlytos naujos idėjos, buvo sudaromi apibūtinimai ir kuriamos hipotezės, naujos teorijos, atliekami eksperimentai ir t. t.

Relevantiškumo problemos iškilimas. Pirmieji relevantiškumą iš nerelevantiškumo pozicijų aptarę informacijos paieškos teorijos pionieriai K. S. Musersas (1950 m.), J. V. Peri (1951 m.) ir M. Taubė (1955 m.). Nerelevantiškumas buvo vadinamas įvairiais terminais: „klaidinga koordinacija“ (false coordination), „triukšmas“ (noise), „neigiamos klasės“ (negative classes) ir kt., o jo priežastimi laikoma kodavimo schemų (coding schemes), indeksavimo kalbų (indexing languages) neefektyvumas, netikslumas. Pažiūroms į relevantiškumą vystyti labai didelės įtakos turėjo K. Šenono informacijos teorijos nagrinėta informacinio triukšmo sąvoka. Į relevantiškumą buvo žiūrima kaip į santykio tarp užklauskos terminų ir dokumento indeksavimo terminų rezultatą. Tai pagimdė tokią sąvoką, kaip „dokumentų giminiškumas“ (relatedness among documents), tuo metu tapusia relevantiškumo sąvokos sinonimu. Relevantiškumas šiuo laikotarpiu buvo

laikomas tik IPS ir joje esančių dokumentų vidinės organizacijos savybe.

Pirmosios diskusijos. Relevantiškumo sąvoka pirmą kartą plačiau nagrinėta 1958 m. Vašingtone įvykusiame tarptautiniame konferencijoje mokslinės informacijos klausimams aptarti. Su šia konferencija siejasi daugelis dabartinių pažiūrų į relevantiškumą. Čia B. K. Vikeris pasiūlė skirti du relevantiškumo aspektus: 1) informacijos vartotojo nustatytą relevantiškumą (user relevance), 2) relevantiškumą atitinkamos temos atžvilgiu (relevance to a subject). G. Bar Hilelis reikalavo, kad būtų daromas skirtumas tarp giminiškumo (relatedness) ir relevantiškumo laipsnio (degree of relevance). R. A. Faithornas teigė, kad relevantiškumas, nustatytas, remiantis dokumento tekstu ir kitu atveju — indeksais išreikštu to dokumento turiniu, — negali būti vienodas. Konferencijoje pirmą kartą oficialiai pripažinta, kad relevantiškumas priklauso ne tik nuo vidinio IPS efektyvumo — jam įtakos turi ir subjektyvūs faktoriai. Tačiau gilesnių diskusijų šia tema neįvyko. Iškeltos pažiūros buvo paremtos daugiau loginėmis ir intuityviomis išvadomis. Išsamiau nei teoriškai, nei eksperimentiškai jos pagrįstos nebuvo.

Relevantiškumo teorijų kūrimas. Pasisėdėjusios diskusijos daugelį tyrinėtojų paskatino į relevantiškumą pažvelgti iš abstrakčiosios pusės: kurti matematiniais, statistiniais, loginiais metodais pagrįstas teorijas.

Vieną pirmųjų stambių teorijų 1960 metais paskelbė M. E. Maronas ir J. L. Kunsas. Jie relevantiškumą apibūdino tikimybių teorijos metodais. Šios teorijos terminais išreiškė santykius, susidarancius tarp informacijos vartotojo, jo užklauskos ir IPS išduoto dokumento, įvedė sąvoką „relevantiškumo kiekis“ (relevance number) — tikimybę, rodančią, kada vartotoją paten-

kins IPS išduoti dokumentai. M. E. Marono ir J. L. Kunso teorija reikšminga tuo, kad joje išskirti relevantiškumui įtakos turintys faktoriai (užklaustos, jų dalykinis turinys, IPS, dokumentai), nustatyta, jog šie faktoriai susiję vienas su kitu, santykiei tarp faktorių apibūdinti tikimybiškai.

1964 m. D. J. Hilmanas paskelbė teoriją, skirtą relevantiškumo sąvokai apibrėžti. Jis teigė, kad užklausų ir dokumentų atitikimas susijęs su sąvokų giminingumo problema. Visiems santykiams, susijusiems su relevantiškumo sąvoka, išreikšti D. J. Hilmanas panaudojo formaliąją logiką ir ypač R. Karnapo sąvokų formavimo teoriją.

Trečiąją teoriją 1964 m. pateikė F. Gofmanas. Jis relevantiškumą apibrėžė kaip dokumento suteiktos informacijos, atsakant į užklausą, matą. V. Gofmanas nurodė matematinėje matų teorijoje keliamas sąlygas, kurias relevantiškumas, būdamas tam tikru matu, privalo patenkinti: 1) jis turi būti išreikštas teigiamu realiu skaičiumi, 2) atskirų dalių išmatavimų suma turi būti lygi dalių sumos išmatavimui, 3) mažesnis dydis privalo turėti mažesnę išmatavimą, ir atvirkščiai, 4) būtinai reikalingas atskaitos taškas — absoliutus nulis. Po to V. Gofmanas įrodė dvi teoremas: 1) jeigu relevantiškumas apibrėžtas tiksliai esamo giminingumo tarp užklaustos ir dokumento pagrindu (be ryšio su kitais dokumentais), tai jis nebepatenkina antrosios sąlygos: pavienių dokumentų relevantiškumų suma gali būti didesnė, negu visų dokumentų, paimtų kartu, relevantiškumas, 2) jeigu relevantiškumas yra apibrėžtas ne tik esamo bendrumo tarp užklaustos ir atskirų dokumentų pagrindu, bet ir šių dokumentų savitarpio santykių pagrindu — tai relevantiškumas kaip matas turi visas reikalaujamas savybes.

Šias apie relevantiškumą kaip apie matą iškeltas mintis 1967 m. V. Gofmanas ir V. A. Njuilas, o 1970 m. — V. Gofmanas apibendrino ir toliau išvystė daug gilesniame ir didesniame darbe — bendroje komunikacijų teorijoje. Jie komunikacijos procesui pritaikė epideminio proceso teoriją (informacijos plitimas tolygus ligų plitimui). Teorijoje centrinė yra efektyvaus kontakto sąvoka. Teoremų serija įrodoma, kokiomis sąlygomis vyksta proceso dinamika. Komunikacijos procese išskiriamas smulkesnis procesas — informacijos paieška. Relevantiškumas apibrėžiamas kaip kontakto, susidarančio perduodant informaciją, efektyvumo matas. Nurodoma, kad relevantiškumas daugiausia matuojamas tikimybiniais metodais. Po to matematiniais apibrėžimais ir teoremomis nurodomos sąlygos, kuriomis atsiskleidžia įvairūs ryšiai tarp dokumentų. V. Gofmanas įrodo, kad relevantiškumas nėra susijęs su viena dokumentų, išduodamų atsakant į užklausą, aibe — galimi keli atsakymų variantai. Dokumentai, iš pradžių buvę nerelevantiški, gali tapti relevantiški atitinkamomis aplinkybėmis. Taigi relevantiškumas yra sąlygiškas santykis.

Nė viena paminėta teorija nebuvo išbaigta, nes apėmė tik atskirus relevantiškumo aspektus ir pateikė skirtingus metodus relevantiškumo savybėms ir jį sudarantiems santykiams apibūdinti. Tačiau jos reikšmingos, kadangi išaiškino ir apibrėžė relevantiškumo sąvoką ir nurodė gaires būsimiems eksperimentams.

Relevantiškumo panaudojimas ir tyrimas IPS eksperimentuose. Vertinimo matų, relevantiškumą laikančių kriterijų, kūrimas. Relevantiškumo problemoms spręsti daugeliu aspektų reikšmingi buvo užsienyje atlikti D. Svon-

sono³, J. Seltone⁴, S. Kleverdone⁵, ISILT (information science index languages test), MEDUSA (MEDLARS on-line search formulation and indexing), UCKIS (Retrieval experiments based on Chemical Abstracts Condensates)⁶, TSRS — „Raduga“, „Lastočka“, „Estafeta“, „Bliznecy“, „Smotr“⁷ ir kiti IPS eksperimentai. Kriterijumi, atspindinčią IPS funkcionavimą, beveik vieningai buvo pasirinktas relevantiškumas ir jo pagrindu sukurta keletas rodiklių IPS techniniam efektyvumui vertinti. A. Michailovas, A. Čiornij, R. Giliarevskis nurodo daugiau negu dešimt rodiklių, kriterijumi laikancių relevantiškumą.⁸ Dar kitus rodiklius savo eksperimentuose naudojo Leningrado valst. N. Krupskajos kultūros institutas.⁹ IPS darbo efektyvumui vertinti skirti rodikliai — tai įvairūs išduotų, neišduotų ar sistemos paieškos masyve esančių relevantiškų ir nerelevantiškų dokumentų santykiai. Visi jie vienu ar kitu būdu išvesti iš masyve esančių relevantiškų dokumentų ir šių dokumentų tarpusavio atitikimo lentelės.

Po to, kai relevantiškumas tapo IPS darbo kriterijumi, iškilo svarbi problema — kas jį turi užfiksuoti. Šeštojo de-

šimtmečio pradžioje tam buvo naudojami dokumentai: remiantis atrinktais dokumentais, sudaromos užklauskos, o pasakui žiūrima, kiek šių atrinktų dokumentų IPS išduos. Pagal išduotų ir atrinktų dokumentų santykį ir buvo sprendžiama apie IPS efektyvumą. Bet daugiausia dokumentų relevantiškumą nustatinėjo žmonės. Tai paskatino tirti psichologines relevantiškumo problemas. Vykstant IPS tyrimams, buvo kuriamos ir tobulinamos relevantiškumo matavimo priemonės: sudarytos įvairios skalės relevantiškumui nustatyti, aktyviai pradėtas naudoti dokumentų randažavimo metodas.

Šiuo relevantiškumo sąvokos vystymosi etapu relevantiškumas buvo laikomas vien to objekto, kuris yra informacijos sklaidimo tikslas, savybe. Be to, šiuo laikotarpiu tolygiomis buvo laikomos dvi skirtingos sąvokos: relevantiškumas (релевантность —, relevance), parodantis informacijos šaltinio ir užklauskos ryšį komunikacijos procese, ir relevantiškumo nustatymas (определение, оценка релевантности, relevance judgement), parodantis tik tam tikrą atsakymą į užklauską. Tai sukėlė daug neaiškumų.

³ Рис А. Оценка информационных служб и информационных систем.— В кн.: Сборник переводов по вопросам информационной теории и практики.— М., 1969, т. 13, с. 84.

⁴ Там же, с. 76—77.

⁵ Сборник переводов по вопросам информационной теории и практики.— М., 1967, т. 9, с. 4—88; М., 1969, т. 12, с. 4—94.

⁶ Jones K., Rijsbergen G. J. Information retrieval test collection.— Journal of Documentation, 1976, v. 32, Nr. 1, p. 59—75.

⁷ Соколов А. В. Влияние субъективных факторов на качество работы информационно-поисковых систем.— Научно-техническая информация. Сер. 2, 1967, № 12, с. 29—36; Соколов А. В. и др. Исследование сравнительной эффективности традиционных и нетрадиционных ИПС.— В кн.: Ленингр. гос. ин-т культуры. Труды. Л., 1970, т. 23, с. 3—35; Соколов А. В. Источник потерь информации в традиционных ИПС и возможности их компенсации (по результ. эксп. исследований).— Научно-техническая информация. Сер. 2, 1967, № 9, с. 26—33; Соркин А. М. Изучение эффективности информационного поиска с помощью картотек, организованных по иерархическим схемам классификации.— В кн.: Ленингр. гос. ин-т культуры. Труды. Л., 1970, т. 23, с. 36—53.

⁸ Михайлов А. И., Чёрный А. И., Гиляревский Р. С. Основы информатики.— М., 1968, с. 306—315.

⁹ Соколов А. В. и др. Исследование..., с. 3—35; Соркин А. М. Изучение..., с. 36—53.

Antrasis diskusijų etapas. Eksperimentiniai IPS tyrimai vėl suaktyvino diskusijas dėl relevantiškumo. 1964 m. įvyko konferencija, kurioje buvo gvil-denama ir keletas su relevantiškumu susijusių klausimų. Antruoju diskusijų etapu iškelti ir aptarti (nors nebūtinai išspręsti) klausimai gali būti suklasifikuoti taip: 1) relevantiškumo sąvokos apibrėžimas, 2) faktoriai, turintys įtakos relevantiškumo nustatymui, 3) relevantiškumo ir relevantiškų dokumentų objektyvumas ir subjektyvumas.

Apibrėžimai. Vienas pirmųjų bandymų paaiškinti reiškinį — tai jo apibrėžimas. Relevantiškumo sąvokos apibrėžimų reikšmė įvairiapusė: 1) apibrėžimas parodo reiškinio esmę, 2) tampa instrukcija, padedanti įvertinti IPS išduotų dokumentų relevantiškumą, 3) apibrėžimas — tai faktorius, turintis įtakos eksperimentų rezultatams. Relevantiškumo apibrėžimų tiek daug, kad A. M. Rysas ir D. G. Sulcas, 1967 m. juos analizuodami, sudarė du sąrašus, nenurodydami nei apibrėžimų autorių, nei šaltinių. Apibrėžimuose sakoma, kad relevantiškumas yra dokumento informatyvumo duotai užklausiai matas; giminytės tarp dokumento ir užklaunos matas; atsakymas, patenkinantis užklausa; atsakymo naudingumo matas ir t. t. Relevantiškumas buvo išreikštas ir algoritmu¹⁰ (žr. schemą Nr. 2). Tarybinėje literatūroje vieną pirmųjų apibrėžimų 1963 m. pateikė M. Kolčinskis. Jis relevantišką vadina tą informaciją, kuri atitinka užklausa¹¹. A. Čiornyj, aptaręs kai kuriuos užsienio mokslin-

kų požiūrius į relevantiškumą, šią sąvoką traktuoja kaip prasmės artumą tarp dokumento ir informacinės užklaunos¹². Vėliau tas pats apibrėžimas pakartotas A. Michailovo, A. Ciornyj, R. Giliarevskio vadovelyje „Informatikos pagrindai“¹³. 1961 m. relevantiškumo apibrėžimą tam tikru aspektu tikslino L. Kozackovas¹⁴. Jis pažymi, kad relevantiškumas organiškai susijęs su mokslinio pažinimo procesu ir atspindi objektyvaus pasaulio ir jo dėsningumų vienybę. Pirminiai šaltiniai, susiję su relevantiškumo nustatymu, yra tiesioginiai moksliniai rezultatai — metodai, modeliai, sąvokos, eksperimentiniai duomenys ir t. t. Būtent jie, šie moksliniai rezultatai, ir yra relevantiški ar nerelevantiški duotai užklausiai, kuri taip pat yra mokslinės ar konstrukcinės veiklos rezultatas. Visi moksliniai rezultatai turi materialią dokumento formą ir šiuo metu relevantiškumą daug kas perkelia nuo mokslinių rezultatų į dokumentą. Tačiau reikia atsiminti, kad dokumentas (forma) yra relevantiškas tik todėl, kad užklausiai yra relevantiška dokumente esanti mokslinė-techninė informacija, o ne atvirkščiai. Kada nors mokslinių rezultatų saugojimo materialioji forma bus ESM atmintis, tačiau niekas tada nekalbės apie mašinos relevantiškumą.

Hipotezės. Kitas žingsnis, aiškinant relevantiškumo prigimtį, — įvairių hipotezių iškėlimas. Jose pagrindinis dėmesys buvo skiriamas relevantiškumo objektyvumo ir subjektyvumo problemoms. Pagrindiniai hipotezės iškėl-

¹⁰ Saracevic T. The concept..., p. 131.

¹¹ Колчинский М. Л. О критериях оценки научно-информационной деятельности. — Научно-техническая информация, 1963, № 12, с. 3—11.

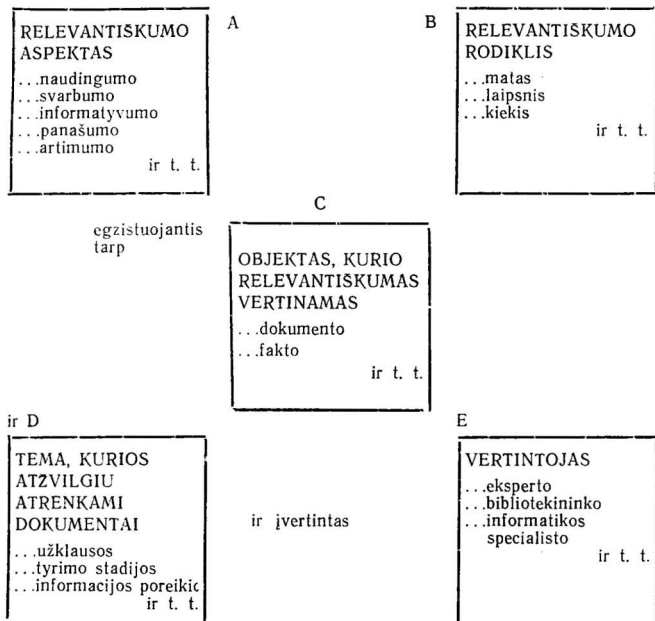
¹² Чёрный А. И. Критерий смыслового соответствия документальной информационно-поисковой системы. — Научно-техническая информация. Сер. 2, 1967, № 9, с. 17—25.

¹³ Михайлов А. И., Чёрный А. И., Гиляревский Р. С. Основы..., с. 285.

¹⁴ Козачков Л. С. О некоторых проблемах релевантности в информатике и науковедении. — Научно-техническая информация. Сер. 2, 1969, № 8, с. 3—11.

RELEVANTIŠKUMO APIBRĖŽIMO ISREIŠKIMAS ALGORITMU

Relevantiškumas yra:



ti teiginiai buvo šie: 1) tiksliai pats informacijos vartotojas gali nustatyti, ar dokumentas relevantiškas, t. y. relevantiškumo įvertinimas gali būti tik subjektyvus; 2) praėjus kuriam laiko tarpui, tas pats informacijos vartotojas to paties dokumento relevantiškumą gali įvertinti skirtingai — tai nulemia naujai įgytos žinios; 3) tų pačių dokumentų relevantiškumas gali būti vertinamas skirtingai ir dėl ieškomos informacijos panaudojimo tikslų įvairumo.

Diskutuojant apie relevantiškumo subjektyvumą ir objektyvumą, susidarė dvi mokyklos, kurioms ryškiausiai atstovavo L. B. Doilis (1963 m.) ir C. A. Kuadra (1964 m.). L. B. Doilis relevantiškumą laikė subjektyvia ypatybe, netinkančia būti jokių matavimo kriterijumi. C. A. Kuadra taip pat pripažino, kad relevantiškumas gali būti subjektyvus, tačiau išvadas daryti siūlė tik tikslų eksperimentinių duomenų pagrindu.

Apie relevantiškumo subjektyvumą ir objektyvumą, apie galimybę relevantiškumą naudoti IPS efektyvumo vertinimo kriterijumi diskutavo ir kiti mokslininkai.

M. Taubė IPS efektyvumo įvertinimą, pagrįstą subjektyviu relevantiškumo kriterijumi, vadino pseudomatematika¹⁵. Dž. Seltonas neigė M. Taubės išvadą¹⁶. Jo nuomone, jeigu relevantiškumą nustato tipiškas informacijos vartotojas, tada apskaičiuotasis IPS efektyvumas — tikrai vertingas rodiklis. Be to, relevantiškus dokumentus gali atrinkti ir kontrolinė informacijos vartotojų grupė. Šiuo atveju relevantiškumo duotai užklausa laikomas toks dokumentas, kuriuo relevantiškumą pripažįsta grupės dauguma. Toks dokumento relevantiškumo įvertinimas, Dž. Seltono nuomone, taip pat patikimas.

1963 m. buvo naujai pažiūrėta į relevantiškumo problemą. A. Sokolovas teigia, kad tai padarė L. Doilis¹⁷. Kitoje literatūroje¹⁸ nurodoma, kad tais metais naujo požiūrio iškėlėjas buvo ir A. M. Rysas bei T. Saracevičius. Minėti tyrinėtojai pareiškė, kad informacijos vartotojo pateikta užklausa ne visada atitinka jo informacijos poreikį, todėl tie dokumentai, kurie relevantiškai užklausiai, nebūtinai relevantiškai informacijos vartotojo poreikiui. Todėl galima skirti du to paties dokumento relevantiškumo aspektus: relevantiškumą užklauskos atžvilgiu ir relevantiškumą informacijos poreikio atžvilgiu. 1966 m. V. Gofmanas ir V. Njuilas dokumentų atitikimą informacijos vartotojo poreikiui pavadino pertinentiškumu (pertinentность — pertinence).

Kitos trys hipotezės atsirado tuo metu vykdomų eksperimentų dėka. Jos turėjo didelę reikšmę, nes apibendrinio, suklasifikavo faktorius, turinčius įtakos dokumentų relevantiškumo nustatymui.

1967 m., remdamiesi eksperimentų rezultatais, G. A. Kuadra ir R. V. Keteris pareiškė, kad relevantiškumo nustatymui įtakos turi penkios pagrindinės klasės: dokumentai, užklauskos, vertinimo sąlygos, vertintojai ir relevantiškumo matavimo būdai (žr. schemą Nr. 3).

Kitą hipotezę 1967 m. iškėlė A. M. Rysas ir D. Šulcas. Jie dokumento relevantiškumo vertinimą pavėdė informacijos vartotojui arba kitam jį pavaduojančiam asmeniui. Buvo teigiama, kad, esant tam tikrom sąlygom, — iki galo suformuluotai užklausiai, žinant informacijos vartotojo visuomenėje atliekamas funkcijas, informacijos panaudojimo tikslą, atsižvelgiant į relevantiškumo vertinimo aplinką ir laiką, — gauti rezultatai yra patikimi ir objektyvūs. Panašias mintis išreiškė ir A. Sokolovas.¹⁹

Dar vieną hipotezę 1967 m. suformulavo J. O'Konoras. Jis didžiausią dėmesį skyrė toms priežastims, dėl kurių kai kurie ekspertai tų pačių dokumentų relevantiškumą įvertina nevienodai. Pagal J. O'Konorą, šios priežastys — tai nevienoda užklauskų ir dokumentų interpretacija. Užklauskos dažnai suformuluojamos neišsamiai, būna dvi-prasmės ir pan. Todėl, J. O'Konoro manymu, daug nesutarimų, vertinant dokumentų užklauskos esmę.

Specializuoti relevantiškumo eksperimentai. Šeštuoju ir septintuoju dešimtmečiais atlikta keletas specialiai rele-

¹⁵ Taube M. A note on the pseudo-mathematics of relevance.— American Documentation, 1968, v. 16, No 2, p. 69—72.

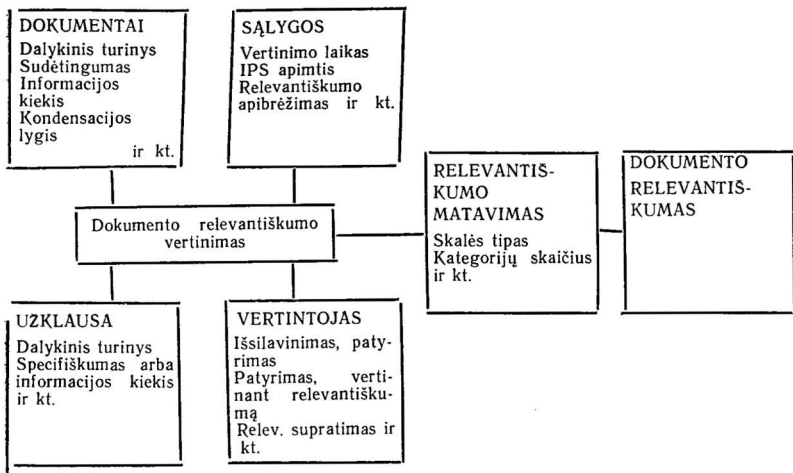
¹⁶ Михайлов А. И., Чёрный А. И., Гиляревский Р. С. Основы..., с. 304—305.

¹⁷ Соколов А. В. Семантические проблемы релевантности.— В кн.: Ленингр. гос. ин-т культуры. Труды. Л., 1970, т. 23, с. 90.

¹⁸ Saracevic T. Relevance.— Journal the American Society for Information Science, 1975, v. 26, Nr. 6, p. 331.

¹⁹ Соколов А. В. Семантические..., с. 96—98.

RELEVANTIŠKUMO VERTINIMUI ĮTAKOS TURINTYS FAKTORIAI



vantiškumui skirtų eksperimentų. Eksperimentų metodologija ir rezultatai buvo plačiai aptarti literatūroje ir sukėlė didelį specialistų susidomėjimą. Eksperimentatoriai orientavosi į psichologinius relevantiškumo klausimus, tyrė faktorius, turinčius įtakos relevantiškumo nustatymui. Tai buvo svarbus supratimo apie relevantiškumą etapas, didelis žingsnis į priekį, aiškinant jo prigimtį. Eksperimentai padėjo atskleisti relevantiškumo kaip kompleksinio reiškinio šaknis, išaiškinti priežastis, turinčias įtakos relevantiškų dokumentų atrankai. Eksperimentų dėka geriau suprasti daugelis klausimų. Gauti rezultatai įdomūs informacijos darbuotojams — teoretikams ir praktikiams.

Pirmieji specializuoti relevantiškumo eksperimentai atlikti 1961 m., o iki 1970 m. įvykdyta per 20 tokių eksperimentų. Didžiausius ir išsamiausius eksperimentus atliko A. M. Rysas ir D. G. Sulcas Case Western Reserve universitete ir C. A. Kuadra bei R. V. Keteris firmos System Development Corporation Laboratorijose. Eksperimentų metu nagrinėtos penkios relevantiškumo vertinimui įtakos turinčių faktorių klasės, kurias nustatė C. A. Kuadra ir R. V. Keteris: dokumentai, užklausos, vertinimo sąlygos ir aplinkybės, matavimo priemonės, vertintojai.

1970 m. visus šiuos eksperimentus apibendrino T. Saracevičius²⁰. Remdamasis eksperimentų rezultatais, jis pa-

²⁰ Saracevic T. Ten years of relevance experimentation — a summary and synthesis of conclusions. — Proceedings of the American Society for Information Science, 1970, v. 7, p. 33—36.

darė keletą išvadų (vėliau atlikti eksperimentai šias išvadas patvirtino). Yra labai didelė tikimybė, teigia T. Saracevičius, kad dokumentai arba objektai, kuriuose užfiksuota informacija — tai faktorius, turintis didžiausią įtaką relevantiškumo nustatymo procesui ir relevantiškumui kaip ryšiui, pasireiškiančiam komunikacijos procesu tarp informacijos sklaidimo šaltinio ir informacijos sklaidimo tikslo. Relevantiškumo nustatymui įtaką daro daugelis dokumento savybių, tačiau svarbiausia jų — dalykinis dokumento turinys. Kuo specifiškesnis dokumento turinys ir daugiau žinoma apie užklausą, tuo labiau tarpusavyje sutariama, vertinant relevantiškumą. Relevantiškumas, nustatytas iš dokumento antraštės, skirsis nuo relevantiškumo, nustatyto iš dokumento teksto, todėl nustatinėti dokumento relevantiškumą iš antraščių reikėtų atsargiai. Taip pat skirsis vieno ir to paties dokumento relevantiškumas, nustatytas iš referato ir iš viso teksto. Šiam skirtumui įtakos turi referato pobūdis, ilgumas, detalumas ir pan. Tarp užklausų, tekstų ir užklausas atitinkančių, t. y. relevantiškų, dokumentų tekste egzistuoja giminingumas, tarpusavio ryšys: jeigu dokumente kuris nors teiginys ar formuluotė atitinka užklausos teiginį ar formuluotę, tai yra labai didelė tikimybė, kad šis dokumentas relevantiškas duotai užklausiai. Giminingumo nepastebima tarp užklausų ir joms nerelevantiškų tekstų. Naudojant įvairių rūšių skales, gaunami truputį skirtingi relevantiškumo įvertinimai. Kuo daugiau skalė turi kategorijų, tuo patogiausiau ja naudotis. Intensyviausiai naudojami kraštutiniai skalių taškai. Dokumentų grupės atitinkamas relevantiškumo dydis gali būti pastovus net tada, kai vertintojai yra skirtingos kvalifikacijos. Dalykinės vertintojo žinios — tai bene svarbiausias faktorius

iš visų kitų žmonių savybių, turinčių įtakos relevantiškų dokumentų atrankai.

Distribucija. Relevantiški dokumentai turi tam tikrus pasiskirstymo dėsnius. Dalis jų buvo tyrinėti relevantiškumui skirtose teorijose ir eksperimentuose. Relevantiškų dokumentų distribucijai nagrinėti taikomi ir kitų informacijos procesų dėsniai, nes jų panaudojimas padeda suprasti ir išspręsti nemažą relevantiškumo problemų²¹.

Iš bibliotekinių praktikos gerai žinoma, kad, atsakant į daugumą informacijos vartotojų užklausų, t. y. atrenkant relevantiškus dokumentus, panaudojama tik nedidelė informacijos masės dalis. Likusioji dalis panaudojama labai retai, kartais ir iš viso nenaudojama. Šį reiškinį 1956 m. tyrė D. J. Urkūhartas, 1964 m. G. Beugeris, V. Gofmanas ir kt. Buvo įtikinamai įrodyta, kad informacijos masę sudarančių dokumentų dažnumas tikrai nėra atsitiktinis — jis turi savo dėsnius.

1949 m. G. Cipfas suformulavo dėsnį, teigiantį, kad išrikiavus visus teksto žodžius mažėjančio dažnumo tvarka ir sudauginus kurio nors žodžio dažnumo dydį su to žodžio eilės numeriu, visada gaunamas pastovus dydis. G. Cipfo dėsnis leidžia giliau pažvelgti į relevantiškų dokumentų pasiskirstymą. Be to, jis labai artimas Bredfordo dėsniui, kuris glaudžiai susijęs su relevantiškumo problema.

S. Bredfordas 1948 m. paskelbė dėsnį: „Jeigu žurnalai išdėstyti straipsnių apie duotą objektą reikšmingumo mažėjimo tvarka, tai jie gali būti suskirstyti į branduolį, kuriame yra daugiausia informacijos apie objektą, ir zonas, turinčias tiek pat straipsnių, kaip ir branduolys; branduolį ir zonas sudarančių žurnalų kiekiai tarpusavyje santykiuoja kaip $1 : n : n^2 : n^3 \dots$ “ S. Bred-

²¹ Saracevic T. The concept..., p. 139—151.

fordo formulę gilino F. Leimkuhleris (1967 m.), J. A. Brukas (1968–69 m.), B. C. Vikeris (1948 m.) ir kt. Bredfordo, Cipfo ir Lotkos dėsnius savo tyrinėjimuose 1968 m. panaudojo L. Kozáčkovas ir L. Chursinas,²² analizuodami mokslinės literatūros augimo kryptis, 1969 m.—L. Kozáčkovas,²³ aptardamas relevantiškumo informacijos augimo, išsisklaidymo, senėjimo dėsnius. 1972 m. V. Maksimovas iškėlė hipotezę, kad daugeliui mokslininkų žurnalai, pagal Bredfordo dėsnį esantys branduolyje, yra neinformatyvūs, nes mokslininkai keičiasi informacija su savo kolegomis individualiai, dar prieš jai pasirodant spaudoje.²⁴

1926 m. A. J. Lotka suformulavo dėsnį, kuriame sakoma, kad n dokumentų sukuriančių žmonių skaičius proporcingas santykiui $1:n^2$. 1963 m. D. J. Praisas pritaikė šią formulę, tyrinėdamas mokslo, mokslinės literatūros, jos autorių didėjimo, literatūros panaudojimo ir visų šių faktorių tarpusavio pasiskirstymo dėsnius. Analizuodamas D. J. Urkusharto rezultatus, J. D. Praisas pastebėjo, kad autorių produktyvumo kreivė tokia pat, kaip ir žurnalų panaudojimo kreivė. Vadinas, autorių produktyvumas turi tuos pačius dėsnius, kaip ir tų autorių sukurtos produkcijos panaudojimas. Tai rodo, kad autorių produktyvumo pasiskirstymas glaudžiai susijęs su relevantiškumo problema.

Dokumente pateikta kito dokumento citata ar bibliografinė nuoroda mažai pasako apie pastarojo dokumento relevantiškumą. Citavimo dėsnius 1956 m. tyrinėjo D. J. Praisas. Jis analizavo

skirtingu laiku išleistų kūrinių citavimo dažnumą ir nustatė, kad dažniausiai cituojami naujausi dokumentai. Kuo senesnis dokumentas, tuo rečiau jis cituojamas. Įvairiu metu išleistų kūrinių citavimo dažnumas sutapo su dokumentų panaudojimo dažnumo ir autorių produktyvumo pasiskirstymo dėsniais.

Kompleksiškai nagrinėjant kuriam nors objektui skirtos literatūros augimą, senėjimą, išsisklaidymą, citavimą ir panaudojimą, išaiškinama tos literatūros struktūra. Šitokio pobūdžio tyrinėjimus atliko V. Gofmanas ir D. Vorensas. Jie analizavo dviem objektams relevantiškų dokumentų, išleistų per šimtmetį, struktūrą. Buvo pastebėta, kad žurnalų ir autorių skaičių santykis visada pastovus. Vadinas, ne straipsnių, o autorių skaičius lemia relevantiškos informacijos išsisklaidymą žurnaluose.

* * *

Relevantiškumo sąvokos supratimas iki aštuntojo dešimtmečio vystėsi sparčiai. Atlikta daug tyrinėjimų, pareikšta vertingų minčių. Septintajame dešimtyje sukurtos gilios, matematiniais metodais pagrįstos relevantiškumo teorijos. Jos aktualios, reikšmingos ir dabartiniu metu. Iki 1970 m. atlikta nemaža relevantiškumo tyrimo eksperimentų. Susidarė keletas požiūrių į relevantiškumą.²⁵

Vienas jų — sisteminis požiūris (system's view of relevance). Šis požiūris atsirado, tik iškilus relevantiškumo problemai. Buvo manoma, kad relevantiškumas daugiausia priklauso nuo IPS vidinės organizacijos lygio.

²² Козачков Л. С., Хурсин Л. А. Модель роста числа научных публикаций на основе закона Лотка—Бредфорда—Ципфа.— Научно-техническая информация. Сер. 2, 1968, № 7, с. 3—8.

²³ Козачков Л. С. О некоторых проблемах релевантности в информатике и науковедении.— Научно-техническая информация. Сер. 2, 1969, № 8, с. 3—11.

²⁴ Максимов В. В. «Мгновенные» классификации наук и проблема релевантности научной информации.— В кн.: Наука и технические вопросы истории и теории. Л., 1972, вып. 7, с. 64—68.

²⁵ Saracevic T. Relevance..., p. 326—339.

Vidiniai IPS aspektai turi įtakos jos veikimui ir tuo pačiu — relevantiškų dokumentų išdavimui. Tai, kaip informacijos šaltinis (IPS) manipuliuoja su informacija, priklauso nuo jo kontakto su informacijos sklaidimo tikslu (informacijos vartotoju ar kitu žmogumi) efektyvumas. Tačiau tai nėra vienintelis faktorius, turintis įtakos relevantiškumui. Todėl sisteminis požiūris į relevantiškumą, nors iš dalies ir teisingas, yra neišbaigtas. Ryškiausias jo trūkumas tas, kad nepripažįstama dokumentų atrankos į IPS svarba relevantiškumui.

Kiek vėliau susiformavo informacijos sklaidimo tikslo požiūris į relevantiškumą (destination view of relevance). Jo šalininkai (C. A. Kuadra, R. V. Keteris ir kt.) pagrindinį dėmesį skyrė relevantiškumo vertinimo procesui, labiausiai domėjosi relevantiškumo vertinimui įtakos turinčiais faktoriais. Vystantis šiam požiūriui, buvo iškelta keletas hipotezių, atlikta eksperimentų. Tačiau negalima teigti, kad hipotezėse ir eksperimentuose nagrinėti visi relevantiškumui įtakos turintys faktoriai. Informacijos sklaidimo tikslo požiūrio atstovai painiojo relevantiškumo ir relevantiškumo vertinimo sąvokas.

Lygia greta su sisteminiu ir informacijos sklaidimo tikslo požiūriais išliko ir dalykinės literatūros požiūris į relevantiškumą (subject literature view of relevance). S. Bredfordas, G. Cipfas, A. J. Lotka, L. Kozačkovas ir kiti mokslininkai tyrinėjo literatūros struktūros dėsnius ir sąsajas, pasireiškiančias tarp literatūros ir užklaudos. Dalykinės literatūros požiūris į relevantiškumą turi didelę svarbą kitų požiūrių formavimuisi ir komunikacijos proceso (ypač to, kuris susijęs su IPS) pažinimui. IPS tikslas — stiprinti kontaktą tarp dalykinės literatūros ir informacijos vartotojo. Todėl literatūros struktūros ir ry-

šio su užklausomis dėsnių pažinimas turi įtakos sisteminio, informacijos sklaidimo tikslo ir kitų požiūrių į relevantiškumą vystymuisi.

Dalykinių žinių požiūris į relevantiškumą (subject know ledge view of relevance) išliko, atskyrus relevantiškumo ir pertinentiškumo sąvokas. Šio požiūrio atstovai nagrinėja santykius tarp žinių apie dalyką ir užklaudos apie šį dalyką. Dalykinės žinios ir dalykinė literatūra tarpusavyje aiškiai susijusios. Bet jos nėra vienas ir tas pats, nes turi savo atskirą struktūrą, dinamiką. Jų išsisklaidymas, asociacijos, organizavimas gali būti skirtingi. Todėl ir relevantiškumas, nustatytas pagal objekto ar subjekto santykį su dalykinėmis žiniomis, gali skirtis nuo relevantiškumo, nustatyto pagal objekto ar subjekto santykį su dalykine literatūra.

Pertinentiškumo požiūris į relevantiškumą (pertinence view of relevance) apima santykius tarp informacijos vartotojo trūkstamų žinių ir dalykinių žinių arba literatūros. Jis nagrinėja žinių trūkumo prigimtį, struktūrą, apimtį, kitimo įtaką relevantiškumui. Pertinentiškumo požiūris į relevantiškumą siejasi su informacijos sklaidimo tikslo požiūriu. Šis nagrinėja visus relevantiškumo vertinimui įtakos turinčius faktorius, tarp jų ir žinių trūkumą. Todėl jis platesnis už pertinentiškumo požiūrį. Tačiau pertinentiškumo požiūris turi tvirtesnį pagrindą: filosofai ir psichologai yra daug tyrę, kaip mes sužinome tai, ką žinome.

Relevantiškumas, būdamas viena pagrindinių informatikos sąvokų, mokslininkų dėmesio rate liko ir toliau. Po 1970 metų išsiplėtojo naujas diskusijų etapas. Šioms diskusijoms labai didelės reikšmės turėjo požiūriai į relevantiškumą, pareikšti septintajame dešimtmetyje.

Резюме

Релевантность — это одно из самых основных понятий в теории информации. В статье обозреваются понятия релевантности до 70-ых годов.

Ранние пионеры информационного поиска заметили, что не вся информация, выдаваемая ИПС, релевантна. Потому проблема релевантности стала актуальной. Начались первые дискуссии. Были выдвинуты теории, в которых применялись математические, логические и другие методы для решения проблемы реле-

вантности. Многочисленными определениями исследователи старались выразить сущность релевантности. В гипотезах обсуждались проблемы субъективности и объективности релевантности, классифицировались факторы, оказывающие влияние на оценку релевантности. Эти факторы исследовались и экспериментально. Много работ было объявлено по дистрибуции релевантных документов. Сформировалось несколько подходов к релевантности.

THE EVOLUTION OF RELEVANCE CONCEPT

REGINA JUREVICIUTE

Summary

Relevance is one of the most fundamental concepts encountered in the theory of information retrieval. The article reviews the evolution of relevance notion till 1970.

The early pioneers of information science quite correctly recognized that not all that would be retrieved would be relevant. So the problem of relevance became urgent and the first challenge began. Theoretical works on relevance have resulted in drawing methods and results from other and more vigorous subjects

to the study of relevance. Numerous definitions of relevance were offered. The hypotheses by large concentrated on enumerating and classifying the factors that affected relevance judgement, on subjectivity and objectivity of relevance. Specialized experiments on relevance were carried out. In bibliometrics some laws have been uncovered, quantitative observations have been made. Some different views on relevance have been formed.