Csabay Károly
Országos Pedagógiai Könyvtár és Múzeum

Horváth Tibornak

„Ott, ott a part!"

...az ingó padlaton, mint öreg tengerész
a hullámok fölött magasba tartanálak,
mutatva azt a fényes csíkot a láthatáron:
ott, ott a part! –
Én már nem érem el,
de látni fogom talán a te szemeddel.
Rónay György: Az öreg tengerész

Teljességre törekvő keresőkérdés – használhatatlanul sok találat. Minuciózusan megfogalmazott keresőkérdés – kevés vagy esetleg nulla találat. Ősrégi dilemma. Van-e arra mód, hogy egy keresőrendszer a megfelelően pontosan megfogalmazott kérdés mellé a „nem túl messze levő" találatokat összegyűjtse? – Válaszunk: van.

Az alábbiakban kísérletet teszünk egy hosszabb távú projekt felvázolására. Célja olyan egzakt fogalmak kialakítása, amelyeknek segítségével a – nevezzük így – *holdudvaros* keresés számítógéppel kezelhetővé válik.

Holdudvarosnak nevezzük a keresést, amikor a keresési pont bizonyos (a kereső által megadott sugarú) *környezetében* minden találat érdekes. Ahhoz, hogy egy sugár mérhető legyen, *távolságfogalmat* kell bevezetni. Úgy képzeljük, hogy első lépésben az atomi fogalmak (értsd: egyszerű tárgyszavak) közötti távolság definiálható kétféle szempontból, majd a belőlük épített struktúrák távolsága, végül a többféle távolság „összeterelése" egyetlen távolsággá. A szóban forgó struktúra egy speciális gráf, a *dag* lesz. Az atomok közötti távolság mérése két tényezőn alapszik: egyrészt a tezaurusz nyújt támpontot két atom távolságára vonatkozólag, másrészt maga a könyvtári anyag, mint statisztikai minta. Mert a mintától – a könyvtártól – független az a tudás, amely azt viszi bele a keresésbe, hogy pl. Kalifornia az Egyesült Államokban van, ugyanakkor a mintából statisztikai úton derül ki például az, hogy Kaliforniában sokat foglalkoznak afáziás gyerekekkel.

A vektoralgebrából ismert *skaláris szorzat* annak kifejezője, hogy „az egyik vektor mekkora árnyékot vet a másikra", azaz hogy a két dolognak mennyi köze van egymáshoz. Ha két vektor ortogonális, merőleges egymásra – ami abban tükröződik, hogy skaláris szorzatuk nulla –, akkor az általuk képviselt két „dolog" (fizikai mennyiség vagy ítélet) független egymástól. Célunk skaláris szorzatot definiálni a fent ismertetett atomok között, mert a skaláris szorzat révén hosszúság-, majd távolságfogalom is adódik. Egy ilyen matematikai apparátus hozzásegítene egyszersmind ahhoz is, hogy mérjük információs tezauruszaink redundanciatartalmát; matematikai nyelven: a tezaurusz minél inkább ortogonális rendszer, annál kevesebb redundanciát hordoz.

A végcél, mely szemünk előtt lebeg, egy olyan könyvtári szolgáltatás, amelyben a könyvtáros a dokumentumokat egy adekvát daggal írja le, majd a könyvtárhasználó az igény megfogalmazásakor egy – vagy több – hasonló dagot alkot meg. (Ebben esetleg egy természetes nyelvi interfész lehet a segítségére.) Ugyancsak a könyvtárhasználó adja meg azt a *tűréshatárt* egy valós szám formájában, amelyet mint környezeti sugarat képzel el megfogalmazott dagja körül: így hozza létre az említett holdudvart. A keresőrendszer pedig rendelkezésére bocsátja a holdudvaron belüli összes találatot.

Nem a fellegekben járunk! Az *Országos Pedagógiai Könyvtár és Múzeum*ban üzemelő *PRECIS* rendszer hasonló struktúrákkal dolgozik; a természetes nyelvi interfész alapmodulja *Prószéky Gábor*nak, az OPKM munkatársának fejlesztése, a magyar morfológiai elemző készen áll. A pedagógiai szakterületen kellően finom távolságfogalom kialakításához szükséges tezaurusz, a *TAPIR* ugyancsak az OPKM könyvtárosainak köszönhetően rendelkezésre áll. Mire várnánk?

A távolság

Matematikai értelemben a távolság a *metrikus terekben* szerepet játszó függvény. Először is tehát definiálnunk kell a metrikus teret:

Egy X halmaz és egy $\delta : X \times X \rightarrow X$ függvény (X, δ) együttesét *metrikus tér*nek nevezzük, ha teljesülnek az alábbiak:

$$\text{(i)}\quad \forall\, x, y \in X : \delta(x, y) \geqslant 0 \;\&\; \delta(x, y) = 0 \Leftrightarrow x = y$$

$$\text{(ii)}\quad \forall\, x, y \in X : \delta(x, y) = \delta(y, x) \tag{1}$$

$$\text{(iii)}\quad \forall\, x, y, z \in X : \delta(x, y) + \delta(z, y) \geqslant \delta(x, y)$$

Szavakkal:

(i) semelyik két pont között nem lehet a távolság negatív, illetve akkor és csak akkor nulla, ha a két pont azonos;

(ii) a távolságfogalom szimmetrikus: x és y között ugyanannyi a távolság, mint y és x között;

(iii) a kerülőút nem lehet rövidebb: ha az x és az y közti útba beiktatunk egy z pontot, a kapott két távolság összege nem lehet kevesebb az eredeti távolságnál.

A fenti kritériumok közül – mint megmutatjuk – (i) enyhítendő oly módon, hogy megengedhető, hogy a nem azonosak között is lehet nulla távolság. Hogy ezt megmutathassuk, be kell vezetnünk egy matematikában szintén mindenütt szerepet játszó fogalmat, az *ekvivalenciareláció* fogalmát.

Egy H halmaz feletti $R \subseteq H \times H$ relációról azt mondjuk, hogy *ekvivalenciareláció,* ha

(R) $\forall x \in H: (x, x) \in R$ *(reflexivitás)*

(S) $\forall x, y \in H: (x, y) \in R \Leftrightarrow (y, x) \in R$ *(szimmetria)* (2)

(T) $\forall x, y, z \in H: \big((x, z) \in R \,\&\, (z, y) \in R\big) \Rightarrow (x, y) \in R$ *(tranzitivitás)*

A halmazt, amely felett definiálva vannak, az ekvivalenciarelációk úgynevezett *ekvivalenciaosztályokra* bontják. Azonos ekvivalenciaosztályba éppen az egymással relációban álló elemek kerülnek. Először is megmutatjuk, hogy ha (1)-ben (i) helyett az alábbi

(i') $\forall x, y \in X: \delta(x, y) \geq 0 \,\&\, \delta(x, x) = 0$

pontot alkalmazzuk, akkor az $(x, y) \in N \Leftrightarrow \delta\ (x, y) = 0$ értelmezéssel definiált $X \times X$-beli N reláció ekvivalenciareláció. Az egyszerűbb írásmód kedvéért az xNy jelölést alkalmazzuk $(x, y) \in N$ helyett:

Tekintsük tehát (2)-t, és:

(R) N eleget tesz a reflexivitás elvének, hiszen (i') miatt xNx

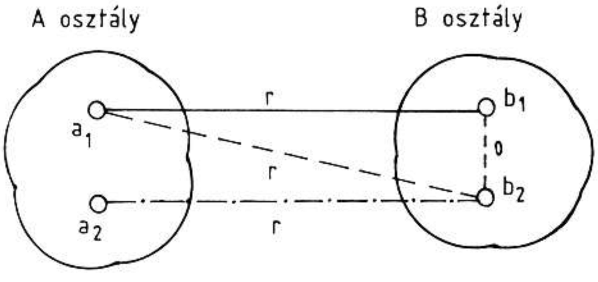
(S) N eleget tesz a szimmetria követelményének, hiszen $yNy \Rightarrow \delta(x, y) = 0 \Rightarrow \delta(y, x) = 0 \Rightarrow yNx$

(T) és végül N tranzitív is, hiszen ha xNz és zNy fönnáll, akkor xNy-nak is fönn kell állnia, tudniillik ha $\delta(x, y)$ nagyobb lenne, mint nulla, az ellentmondana a „kerülőút nem lehet rövidebb" elvnek – ekkor ugyanis z-n keresztül nulla távolság adódna x és y között.

Ezzel bebizonyítottuk, hogy a „nulla távolságra van" fogalma csakugyan ekvivalenciareláció, ami azt eredményezi, hogy az egymástól nulla távolságra levő pontok ebben az enyhített metrikus térben ekvivalenciaosztályokat alkotnak. Most már csak annyi van hátra, hogy bevezessük a két osztály közötti távolság fogalmát. Azt állítjuk, hogy *közömbös, hogy kik a reprezentánsok* az egyes osztályokból A és B osztály távolsága egy tetszőleges $a \in A$, illetve $b \in B$ elem közötti távolság, bárhogyan választjuk is a és b elemeket.

Bizonyításul legyen a_1, $a_2 \in A$, valamint b_1, $b_2 \in B$. Legyen továbbá $\delta(a_1, b_1) = r$. Állítjuk, hogy ekkor $\delta(a_2, b_2)$ is r. Tekintsük ugyanis például $\delta(a_1, b_2)$ távolságot. Ennek r-nek kell lennie, tudniillik, ha akár hosszabb, akár rövidebb lenne r-nél, a_1, b_1, b_2 pontok megsértenék a „kerülőút nem lehet rövidebb" elvet. De ugyanilyen megfontolásból adódik, hogy $\delta(a_2, b_2) = r$, különben a_1, a_2, b_2 pontok szolgáltatnának példát az elv sérelmére *(1. ábra).*

A távolságfogalomnak ilyen enyhítésére azért volt feltétlenül szükség, hogy ne okozzanak gondot a továbbiakban olyan esetek, amelyekben a metrikus tér két pontja között a távolság nulla, noha a pontok arisztotelészi értelemben nem *azonosak.* Célunk ugyanis, hogy a most felvázolandó rendszerrel *tárgyszavak,* illetve speciális *tárgyszóstruktúrák* közötti távolságfogalmat alkossunk meg. Nyilván előfordulnak majd olyan tárgyszópárok (kutya / eb), amelyek között nulla távolságot kívánunk értelmezni. Modellünkben az ilyen csoportok fogják a fönt ekvivalenciaosztályoknak nevezett osztályokat alkotni.



1. ábra **Az *A* és a *B* osztály távolsága**

Távolság a tezauruszban

Azt, hogy az általunk használt tezaurusz mennyire lesz speciális gráf (például fa vagy *dag* – azt, hogy mi a dag, lásd később), nem tudjuk megmondani. Éppen ezért most olyan távolságfogalommal kell beérnünk, amely tetszőleges (irányított) gráfban alkalmazható. A kutatás előtt, persze, nyitva áll egy finomabb definíció lehetősége, amennyiben a tezaurusz részéről valamilyen matematikai feltételt garantálni tudunk.

Most két csomópont (tárgyszó) közötti távolság definiálására a (valamely gráfkereső algoritmussal megtalált) minimális út hosszát választjuk, ha ilyen út létezik. Az út hossza kifejezés az úton fekvő élek súlyának összegét jelenti. Hogy a választott definíció metrikát ad, a definícióból következik. Talán annyi kiegészítést érdemes még tennünk, hogy a nem létező összeköttetéseket végtelen nagy távolságként értelmezzük, így az értelmezési tartomány teljes lesz.

Most bevezetett távolságfogalmunk egy általános tudásbázishoz kötődik, melyet a rendszer a tezauruszban tükröztet. Tovább egy olyan irányba kell lépnünk, melyet a kaliforniai afáziás gyerekekkel fémjeleztünk. Az alább bemutatandó matematikai apparátussal az a célunk, hogy a mintából származó összefüggéseket is szerepeltessük a távolságfogalomban.

A skaláris szorzat és a norma

Egy halmazt, amelynek elemeire összeadás és számmal szorzás van értelmezve (pontos definíció lineáris algebrai tankönyvekben található), *vektortérnek* nevezzük. Egy vektortér fölött értelmezhető a *skaláris szorzat* nevű művelet, mely a vektortér két eleméhez hozzárendel egy (valós) számot, s melytől megköveteljük a következő öt tulajdonságot:

Def.: Ha x, y, z egy vektortér elemei, λ (valós) szám, akkor az $\langle x, y \rangle$-nal jelölt *skaláris szorzatra:*

(i) $\langle x, y \rangle = \langle y, x \rangle$ (3)

(ii) $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$

(iii) $\langle x, x\rangle \geqslant 0$ (3)

(iv) $\langle x, x\rangle = 0 \Rightarrow x = \underline{0}$

(v) $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$

Újabb fontos definíció következik:

Def.: Ha *x, y* egy vektortér elemei, λ (valós) szám, akkor az $\|x\|$-val jelölt és *normának* nevezett (valós) számra álljanak fönt:

(i) $\|x\| \geqslant 0$

(ii) $\|x\| = 0 \Rightarrow x = \underline{0}$

(iii) $\|\lambda x\| = |\lambda|\,\|x\|$ (4)

(iv) $\|x+y\| \leqslant \|x\| + \|y\|$

Ismeretes a tétel (pl. [3] 76. old.), hogy ha $\langle x, y\rangle$ skaláris szorzat, akkor az $\|x\| = (\langle x, x\rangle)^{1/2}$ definícióval megadott függvény norma. (Ún. skaláris szorzatból származtatott norma.) Ismeretes továbbá a másik tétel, mely szerint, ha egy norma adott, akkor a $\delta(x, y) = \|x-y\|$ definícióval megadott függvénytávolság (metrika) – (ugyancsak [3]). A fentiek értelmében, ha egy halmaz elemein sikerül skaláris szorzatot definiálni, akkor ebből távolságfogalom adódik.

Tekintsük evégből a szóban forgó könyvtári osztályozási rendszer tárgyszavait *valószínűségi változóknak* abban az értelemben, hogy például egy indexelő könyvtáros minden egyes tárgyszóval „végigvonul" a könyvtári állomány fölött, és minden egyes tételnél egy 0-tól 5-ig (vagy akár 0-tól 100-ig) terjedő osztályzattal kifejezi, hogy a tárgyszó mily mértékben illeszkedik a könyvre. Ekképpen minden tárgyszó egy egyenletes eloszlású valószínűségi változónak felel meg, melynek értékei a könyvtáros által mondott számok, várható értéke pedig ezen számok számtani közepe.

A további vizsgálódás céljából bevezetünk egy ekvivalenciarelációt ezen valószínűségi változók között:

Ekvivalensnek tekintünk két valószínűségi változót, ha a különbségük konstans. Azt, hogy ez a reláció tényleg ekvivalenciareláció, az olvasó önállóan is igazolhatja. Másrészt nyilván értelmes is a definíció: az, ha a két tárgyszóhoz rendelt osztályzatok *minden* könyvön ugyanazzal a konstans mennyiséggel térnek el, nem lehet véletlen, ilyenkor a két tárgyszó ekvivalens.

Valószínűségi változók együttmozgásának, összefüggésének vizsgálatára sok mértéket dolgoztak ki. Egyikük a *kovariancia.* Ha ξ és η két valószínűségi változó, és ξ várható értékét $M(\xi)$-vel jelöljük, akkor kovarianciájuk a

$$cov(\xi, \eta) = M\big((\xi - M(\xi))\,(\eta - M(\eta))\big)$$

mennyiség. Számunkra azért fontos, mert a fent bevezetett ekvivalenciaosztályokra a kovariancia skaláris szorzatként működik.

Bizonyításul tekintsük át (3) pontjait. (i) nyilvánvalóan adódik a kovariancia definíciójából, (ii) belátható a várható érték $M(a\xi) = aM(\xi)$ tulajdonságára támaszkodva. (iii) adódik abból, hogy egy valószínűségi változónak önmagával vett kovarianciája éppen a szórásnégyzete, ami biztosan nem negatív. Ismeretes a $D^2(\xi)$ szórásnégyzet azon tulajdonsága, hogy ha zérus, ebből az következik, hogy ξ (1 valószínűséggel) konstans. A fenti ekvivalenciareláció osztályozása szerinti osztályok között éppen a konstans osztály a zéruselem. Ebből következik (iv). Végül ahhoz, hogy (v)-öt belássuk, $cov(\xi, \eta)$-t kicsit átalakítjuk. Figyelembe véve, hogy a várható érték additív és lineáris, valamint hogy egy várható érték maga konstans mennyiség, azt kapjuk, hogy

$$\begin{aligned} cov(\xi, \eta) &= M\big((\xi - M(\xi))\,(\eta - M(\eta))\big) = \\ &= M\big(\xi\eta - M(\xi)\eta - M(\eta)\xi + M(\xi)M(\eta)\big) = \\ &= M(\xi\eta) - M(\xi)M(\eta) - M(\eta)M(\xi) + M(\eta)M(\xi) = \\ &= M(\xi\eta) - M(\xi)M(\eta) \end{aligned}$$

(v)-öt, a disztributivitást ezzel a formával vizsgáljuk meg:

$$\begin{aligned} cov(\xi, \eta + \zeta) &= M(\xi(\eta + \zeta)) - M(\xi)M(\eta + \zeta) = \\ &= M(\xi\eta) + M(\xi\zeta) - M(\xi)M(\eta) - M(\xi)M(\zeta) \end{aligned}$$

A második sorban éppen $cov(\xi, \eta)$ és $cov(\xi, \zeta)$ összegét ismerhetjük fel.

Ezzel bizonyítást nyert, hogy a kovariancia skaláris szorzatként működik, miáltal a valószínűségi változók (tárgyszavak) felett távolságfogalmat indukál.

Az eddig elmondottakat illusztrálja most egy szerény példa. Az alábbiakban felsorolunk 23 könyvet, melyekhez „tárgyszavak" kapcsolódnak – zárójelben a tárgyszavakhoz tartozó osztályzatokkal.

Nagy László: Tündérkert fejedelme – Báthory Gábor
XVII. sz. 1. fele (4) Báthory Gábor (5) Bethlen Gábor (1) Bocskai István (1) Erdély (2) Erdély története (3) Erdélyi Fejedelemség (4) erdélyi fejedelmek (2) Homonnai Drugeth Bálint (1) Illésházy István (1) Imreffi János (1) Magyarország története (2) Rákóczi Zsigmond (1) Báthoryak (4)

Bitskey István: Pázmány Péter
I. Rákóczi György (1) II. Ferdinánd (1) VIII. Orbán pápa (1) XVII. sz. 1. fele (4) Alvinczi Péter (3) Bethlen Gábor (1) ellenreformáció (3) Erdély története (2) Erdélyi Fejedelemség (1) Forgách Ferenc (2) kassai vértanúk (1) Kempis Tamás (1) Magyari István (1) Magyarország története (1) Pázmány Péter (5) Péchi Simon (1) reformáció (3)

Nagy László: Az erős fekete bég – Nádasdy Ferenc
XVI. sz. 2. fele (4) XVII. sz. 1. fele (4) Báthory Anna (1) Báthory Erzsébet (4) Báthory György (1) Báthory István (1) Báthory Zsigmond (1) Báthoryak (3) Erdély (2) Erdély története (3) Erdélyi Fejedelemség (3) gyurgyevói ütközet (3) Kanizsai Orsolya (1) Magyari István (1) Magyarország története (1) mezőkeresztesi csata (1) Nádasdy Ferenc (5) Zrínyi György (1)

Csonka Ferenc–Szakály Ferenc:
Bocskai kíséretében a Rákosmezőn
XVI. sz. 2. fele (4) XVII. sz. 1. fele (4) Alvinczi Péter (4) Bocatius János (4) Bocskai István (5) Dengeleghy Mihály (2) Dessewffy János (2) Egri István (2) Gáczi András (2) hajdúk (4) hajdúkapitányok (4) Homonnai Drugeth Bálint (4) Illésházy István (2) Kollonich Siegfried (1) Kopcsa Miklós (2) Lalla Mehmed nagyvezír (5) Palotai Mihály (2) Péchi Simon (2) Rákóczi Erzsébet (1) Rákóczi János (1) Rákóczi Lajos (1) Rákóczi Zsigmond (2) Rákócziak (1) Rhédei Ferenc (2) Somogyi György (2) Szabó Lukács (2) Széchy György (2) Székely Mózes (1) Szilasi János (2) Török Bálint (1) Báthory Gábor (1) Báthory István (1) Báthory Zsigmond (1) Báthoryak (1) Báthory Pál (1) Érsekújvár ostroma (4)

Gerendás Lajos & al.:
Görgey Artúr élete és működése Magyarországon
XIX. sz. (3) 1848/49. évi szabadságharc (4) Görgey Artúr (5) Kossuth Lajos (2) Magyarország története (3)

Pusztaszeri László: Görgey Artúr a szabadságharcban
XIX. sz. (4) XIX. sz. 1. fele (4) 1848/49. évi szabadságharc (4) Görgey Artúr (5) Magyarország története (2)

Supka Géza: 1848–1849
XIX. sz. (3) XIX. sz. 1. fele (4) 1848/49. évi szabadságharc (5) Batthyány-kormány (1) Batthyány Lajos (1) Bem József (4) Erdély (1) Erdély története (1) Görgey Artúr (2) Habsburg-ház (1) Jellačić, Joseph (2) Kossuth Lajos (4) Lamberg Ferenc (3) Latour, Theodor (3) Mészáros Lázár (2) Perczel Mór (4) Petőfi Sándor (2) Széchenyi István (2) Szemere Bertalan (2) tizenkét pont (3) trónfosztás (2) választójog (2)

Nemeskürty István:
„Kik érted haltak, szent Világszabadság!"
XIX. sz. (3) XIX. sz. 1. fele (4) 1848/49. évi szabadságharc (5) Aulich Lajos (4) Batthyány Lajos (4) Bem József (3) császári és királyi hadsereg (3) Czetz János (3) Damjanich János (4) Dessewffy Arisztid (4) Görgey Artúr (3) Haynau, Julius Jakob (3) Hentzi, Heinrich (3) Jellačić, Joseph (2) Kiss Ernő (4) Klapka György (3) Knézich Károly (4) Kossuth Lajos (3) Lahner György (4) Latour, Theodor (2) Lázár György (1) Leiningen–Westerburg Károly (4) Mészáros Lázár (4) Móga János (3) Nagy-Sándor József (4) Perczel Mór (2) Pöltenberg Ernő (4) Schweidel József (4) Szemere Bertalan (2) Török Ignác (4) trónfosztás (1) Vécsey Károly (4) Vetter Antal (1) Windisch-Grätz, Alfred (2)

Bethlen István emlékirata – 1944
IV. Károly (2) XX. sz. (3) XX. sz. 1. fele (4) Antibolsevista Comité (4) Bárdossy László (2) Bethlen István (5) bethleni konszolidáció (4) budaörsi csata (1) Dálnoki Veress Lajos (2) Darányi Kálmán (2) Erdély (1) Erdély története (1) Észak-Erdély (1) eucharisztikus kongresszus (1) Gömbös Gyula (3) Habsburg-ház (1) Habsburg-restauráció (1) hadsereg konszolidációja (2) Hitler, Adolf (1) Horthy Miklós (3) Kállay Miklós (3) Károlyi Gyula (3) Károlyi Mihály (3) kassai bombázás (1) Keresztény Nemzeti Egyesülés P. (2) királypuccs (1) Kisgazdapárt (1) konszolidáció (4) Magyar Tanácsköztársaság (1) Magyarország miniszterelnökei (1) Magyarország története (2) Mussolini, Benito (1) nemzetiszocialista mozgalom (1) Népszövetség (1) numerus clausus (1) „Őszirózsás" forradalom (1) Pacelli, Eugenio (1) pénzügyi konszolidáció (4) revízió (1) soproni népszavazás (1) szegedi kormányok (1) szociáldemokrácia (1) Sztálin, Joszif V. (1) Sztójai Döme (1) Teleki Pál (3) trianoni békeszerződés (2) választójog (1) Veesenmayer, Edmund (1) vitézi rend (1) zsidótörvények (1) zsidóüldözés (1)

Nemeskürty István: Requiem egy hadseregért
II. magyar hadsereg (5) II. világháború (3) XX. sz. (3) XX. sz. 1. fele (4) Bárdossy László (1) Hitler, Adolf (1) Horthy Miklós (1) Jány Gusztáv (4) Kállay Miklós (1) Kovács Gyula (3) Magyarország története (2) Nagy Vilmos (3) Stomm Marcell (3) Szálasi Ferenc (1) sztálingrádi csata (3) Weichs, Maximilian (1) Werth Henrik (1) Witzleben, Hermann (1)

Gosztonyi Péter: A kormányzó, Horthy Miklós
I. bécsi döntés (3) II. bécsi döntés (3) IV. Károly (2) Bárdossy László (1) bethleni konszolidáció (1) Ferenc Ferdinánd (1) Gömbös Gyula (2) Habsburg-ház (1) Habsburg-restauráció (1) hadsereg konszolidációja (1) Hitler, Adolf (1) Horthy Miklós (5) Imrédy Béla (3) Kállay Miklós (3) kassai bombázás (1) királypuccs (1) Lakatos Géza (2) Magyar Tanácsköztársaság (1) Magyarország története (1) Népszövetség (1) Sztójay Döme (1) Teleki Pál (3) trianoni békeszerződés (3) trónfosztás (1) választójog (1) vitézi rend (3) zsidótörvények (1) zsidóüldözés (1)

Romsics Ignác: Ellenforradalom és konszolidáció
IV. Károly (2) Ábrahám Dezső (1) Antibolsevista Comité (1) Bethlen István (4) bethleni konszolidáció (5) földreform (2) Friedrich István (3) gazdasági konszolidáció (4) Habsburg-ház (1) Habsburg-restauráció (1) hadsereg konszolidációja (3) Herczeg Ferenc (1) Horthy Miklós (4) Huszár Károly (1) Károlyi Gyula (1) Keresztény Nemzeti Egyesülés P. (1) királypuccs (2) Kisgazdapárt (1) Magyar Tanácsköztársaság (3) Magyarország története (1) Peidl Gyula (1) pénzügyi konszolidáció (3) Peyer Károly (1) revízió (1) Simonyi-Semadam Sándor (1) szegedi kormányok (1) szociáldemokrácia (1) Teleki Pál (1) trianoni békeszerződés (2) trónfosztás (1) választójog (2)

Dombrády Lóránd: A legfőbb hadúr és hadserege
I. bécsi döntés (1) II. bécsi döntés (1) II. világháború (1) IV. Károly (2) XX. sz. (3) XX. sz. 1. fele (3) Bárdossy László (1) Bethlen István (1) budaörsi csata (2) Észak-Erdély (1) földreform (1) Habsburg-restauráció (1) hadsereg konszolidációja (4) Hitler, Adolf (1) Horthy Miklós (5) Imrédy Béla (1) Kállay Miklós (1) királypuccs (2) konszolidáció (2) revízió (1) szegedi kormányok (1) Teleki Pál (1) trianoni békeszerződés (1) trónfosztás (1)

Mayeda, Wataru: Alkalmazott gráfelmélet
alapkör-rendszer (1) alapvágat-mátrix (1) illeszkedési mátrix (1) irányított él (1) irányított kör (1) körmátrix (1) nyílt élsorozat (1) referenciapont (1) teljes illeszkedési mátrix (1) teljes körmátrix (1) teljes vágatmárix (1) fa (3) feszítő részfa (3) gráfelmélet (5)

Lagge, David–Barber, Paul: Információ és készség
információelmélet (1) információ (5) készség (5) pszichológia (4) viselkedés (3) motorikus képesség (2)

Dr. Szász Gábor: Hálóelmélet
gráfelmélet (4) háló (4) hálóelmélet (5) részben rendezett halmaz (4) korlátos háló (2) komplementumos háló (2) disztributív háló (2) moduláris háló (2) osztályozásháló (2)

Rényi Alfréd: Napló az információelméletről
csoportelmélet (1) információelmélet (5) játékelmélet (2) fa (3) feszítő részfa (3) gráfelmélet (3) információ (4) valószínűségszámítás (1) vegyészeti alkalmazás (1)

Shannon, Claude E.–Weaver, Warren:
A kommunikáció matematikai elmélete
információelmélet (5) információ (4) csatorna (3) zaj (3) diszkrét információ (2) folytonos információ (2) Markov-folyamat (1) gráf (1) bizonytalanság (1) entrópia (2) kapacitás (1) kódolás (1)

Frege, Gottlob: Logika, szemantika, matematika
információ (1) logika (5) szemantika (5) matematika (5) nyeylv (2) természetes nyelv (3) formális nyelv (1) jel (1) jelentés (1) számosság (3)

Szabó Árpád: A görög matematika kibontakozása
arányok (3) bizonyításelmélet (1) Eukleidész (4) görög matematika (5) irracionális számok (2) Püthagorasz (3) racionális számok (2) Thalész (3)

Freud Róbert (ed.):
Nagy Pillanatok a matematika történetében
Abel, Niels (1) Appendix (4) arányok (1) bizonyításelmélet (1) Bolyai Farkas (1) Bolyai János (2) Bolyaiak (2) csoportelmélet (1) differenciálszámítás (1) Euler, Leonhard (2) Fermat, Pierre (2) Galois, Évariste (2) görög matematika (2) információelmélet (2) integrálszámítás (2) játékelmélet (2) püthagoraszi iskola (2) Püthagorasz (2) Tentamen (3)

Sain Márton: Matematikatörténeti ABC
Abel, Niels (1) Appendix (1) Arkhimédész (1) Bolyai Farkas (1) Bolyai János (1) Bolyaiak (1) Cantor, Georg F. (1) Cauchy, Augustin (1) Csebisev, Pafnutyij L. (1) D'Alembert, Jean (1) Descartes, René (1) Diophantosz (1) egyiptomi matematika (2) Erdős Pál (1) Eukleidész (1) Euler, Leonhard (1) Fermat, Pierre (1) Fourier, Jean B. J. (1) Galilei, Galileo (1) Galois, Évariste (1) Gauss, Carl F. (1) görög matematika (2) Hilbert, David (1) Kepler, Johann (1) Lagrange, Joseph L. (1) Laplace, Pierre S. (1) Legendre, Adrien M. (1) Leibniz, Gottfried W. (1) Lobacsevszkij, Nyikolaj I. (1) matematikatörténet (5) Monge, Gaspard (1) Möbius, August F. (1) Neumann János (1) Newton, Isaac (1) Pascal, Blaise (1) püthagoraszi iskola (2) Poincaré, J. Henri (1) Püthagorasz (1) Riemann, G. F. Bernhard (1) Tentamen (1) Thalész (1) Weierstrass, Karl (1)

Ribnyikov, K. A.: A matematika története
XVII. sz. (3) XIX. sz. (4) differenciálszámítás (3) Eukleidész (2) geometria (2) görög matematika (2) integrálszámítás (2) infinitézimális módszerek (4) analízis (4) középkor (4) XVIII. sz. (4)

A példa nagyon szegényes, a tárgyszavazás elképesztően igénytelen, a kapott matematikai eredmények mégis meglepően jók. Nevezetesen: a 23 könyvet 301 tárgyszó felhasználásával írtuk le – némelyik tárgyszó nem vesz részt a leírásokban, de a legtöbb igen –; ez azt jelenti, hogy a keletkező kovarianciamátrix 301 × 301 méretű. Ennek bemutatására terjedelmi okból nincs lehetőség. A mátrixban a kovarianciák –0.5 és 2.4 között mozognak. Egy 0.9 fölötti érték már erős pozitív kovarianciának minősül, egy –0.1 alatti értéket pedig erős negatív kovarianciának lehet tekinteni. A 0.002 alatti abszolút értékű kovarianciák azt jelentik, hogy a szóban forgó tárgyszavaknak semmi közük sincs egymáshoz. (Ez utóbbi kifejezésre a példák kapcsán még visszatérünk.)

A fenti 23 kötetes „könyvtár"-ban a legerősebb (2.3 fölötti) kovarianciát az **1848/49. évi szabadságharc** és a **Görgey Artúr** tárgyszavak mutatták. Hasonlóan magas (1.5 fölötti) értéket mutatnak még a következő párok: XIX. sz. 1. fele–1848/49. évi szabadságharc, XIX. sz.–1848/49. évi szabadságharc, XIX. sz.–Görgey Artúr, Bethlen István–bethleni konszolidáció. (Helyes és konzekvens, ha a könyvtáros úgy dolgozik, hogy a XIX. sz. 1. fele és a XIX. sz. összetartozását nem a kovarianciára, hanem a tezauruszra bízza.) Ugyancsak erős kovarianca (1-nél nagyobb) lépett fel a logika, szemantika és matematika tárgyszavak között. Ez bizonyos értelemben a „könyvtár" hibája: túl jó osztályzatokat adtunk nekik egyetlen könyvön (mindhá-romnak ötös), míg egyáltalán nem szerepeltettük őket más könyvek esetében. Ez a helyzet túlértékelte összetartozásukat.

Erős negatív kovarianciát mutat a XVII. sz. 1. fele tárgyszó a XIX. sz., 1848/49. évi szabadságharc, Görgey Artúr tárgyszavakkal (–0.4 alatti értékek). Szintén erős a negatív kovariancia, amikor egyik oldalon Teleki Pál vagy a trianoni békeszerződés, a másikon az információelmélet vagy a gráfelmélet áll (–0.1 alatti értékek). Valószínűleg nem meglepő, hogy a számottevő negatív kovariancia Schweidel József, Szemere Bertalan, Török Ignác, Vécsey Károly, a trónfosztás, a választójog, a vitézi rend vagy akár Érsekújvár ostroma és az információelmélet között szintúgy megjelenik. De mutatkozik ez a negatív kovariancia a Tentamen, Thalész, az infinitézimális módszerek, sőt, az analízis és az információelmélet között is!

Lényegében ortogonálisak (0.002-nél kisebb abszolút értékű kovarianciával) az alábbi párok: I. Rákóczi György–Ábrahám Dezső, II. Ferdinánd–Arkhimédész, a matematikus Riemann–Székely Mózes, erdélyi fejedelem.

Vagyis: ahol az 1848/49-es szabadságharcról van szó, ott Görgey fölbukkanása várható. Ahol Teleki Pálról olvasunk, meg lehetünk győződve róla, hogy nem lesz információelmélet. Ezek a nagy abszolút értékű kovarianciák: valamilyen irányú következtetést lehetővé tesznek. De amikor egy könyvben I. Rákóczi György a tárgy, legalábbis ennek a „könyvtár"-nak az anyagából nem tudunk semmit mondani arról, hogy lesz-e szó benne Ábrahám Dezső miniszterelnökről. *Nekik tényleg semmi közük egymáshoz.*

Nincs mód, és nem is cél, hogy a szerény állomány esetleg a kovarianciák segítségével tükrözzön olyan ismereteket, melyek egy tezauruszban elhelyezhetők: pl., hogy a kassai vértanúk Kőrösi Márk, Grodecz Menyhért és Pongrácz István voltak. Az alábbiakban következő gráfelméleti eszközökkel igyekszünk rávilágítani arra, hogy egy ilyen jellegű „átlépés" növeli ugyan a távolságot a keresőkérdés és a dokumentum között, de megszünteti az áthidalhatatlanságot. Elegendően nagy tűréshatárral az egyetemes történelmen keresztül történelemből kiindulva akár a matematikatörténet is elérhető.

Szintén a helyszűke miatt nem tértünk ki névváltozatokra (Görgei/Görgey), melyeknek kezelését ugyancsak a tezauruszra kell bízni.

A dagok

A továbbiakban mindig *gráf*ot mondunk, és az *irányított* szót elhagyjuk: gráfon mindig irányított gráfot értünk. Azt mondjuk, hogy G egy gráf, amikor G részhalmaza az $L \times V \times V$ Descartes-szorzatnak, ahol L a *címkék,* míg V az *értékek* halmaza. A címke akkor válik *éllé,* amikor egy gráfban felhasználásra kerül, hasonlóan ilyenkor mondjuk, hogy az illető érték *csomópont.* A tényt, hogy

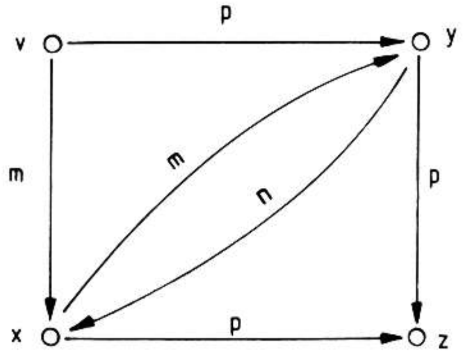
$$(e, n_1, n_2) \in L \times V \times V$$

úgy fogjuk jelölni, hogy $\langle e, n_1, n_2\rangle$. Szavakkal: az *e* él az n_1 csomóponttól az n_2 csomópontig vezet.

Tegyük fel, hogy $V = \{v, w, x, y, z\}$ egy adott értékhalmaz. Hasonlóan $L = \{l, m, n, o, p\}$ a címkék egy adott halmaza. Más szavakkal: adva van öt érték és öt címke. Amint gráfot építünk belőlük az alábbiak szerint:

$$\langle m, v, x\rangle, \langle p, v, y\rangle, \langle m, x, y\rangle, \langle p, x, z\rangle, \langle n, y, x\rangle, \langle p, y, z\rangle. \quad (5)$$

megalkottuk az $N = \{v, x, y, z\}$ csomóponthalmazt és az $E = \{m, n, p\}$ élhalmazt. *N* elemszáma négy, *E*-é három *(2. ábra)*.



2. ábra **Négy csomópontból és háromféle élből szerkesztett gráf**

Be kell vezetnünk egy nagyon fontos fogalmat, az *ős* fogalmát.

Egy *n* csomópont *G*-beli őseinek jelölésére az *Anc(n, G)* szimbólumot alkalmazzuk. Be kell még vezetnünk az $a_k(n, G)$ segédszimbólumot – szavakkal: az *n* csomópont azon őseinek halmazát jelöljük így, melyeknek távolsága *n*-től nem nagyobb, mint *k*. Egzakt módon:

Def.: $$a_1(n, G) = \{n' \mid n' \in N, \exists e \in E: \langle e, n', n\rangle\}$$

$$a_{k+1}(n, G) = \bigcup_{n' \in a_k(n, G)} a_1(n', G)$$

Most $a_k(n, G)$ jelölés felhasználásával definiálhatjuk *Anc(n,G)*-t:

Def.: $$Anc(n, G) = \bigcup_{k=1}^{\infty} a_k(n, G)$$

Nyilvánvaló, hogy *N* végességéből *Anc(n,G)* végessége is következik. Hogy áttekinthetőbbé tegyük a mondottakat, felsoroljuk az (5) gráfban előforduló csomópontok őseit:

$$Anc(v, G) = \emptyset$$

$$Anc(x, G) = \{v, y, x\}$$

$$Anc(y, G) = \{v, x, y\}$$

$$Anc(z, G) = \{x, y, v\}$$

Ha található a gráfban olyan csomópont, amely szerepel saját ősei közt – formálisan, ha

$$\exists n \in N: n \in Anc(n, G)$$

akkor azt mondjuk, hogy *G ciklikus*.

Elérkeztünk a *dag* definiálásához. (A dag betűszó, mely az angol *Directed Acyclic Graph* kifejezésből ered. Voltaképpen nem minden *irányított körmentes gráf* dag, de erre most nem térünk ki.) A dag egy speciális gráf, amelynek van egy kitüntetett csomópontja, a gyökér, melyet általában ϱ-val jelölünk. Egy dag azonosítása során szükség lehet a csomópontok és az élek megjelölésére is, vagyis egy dag jelölésének teljes formája: $\langle \varrho, N, E\rangle$.

Def.: A $\Delta_{L,V}$ halmaz, mint az *L* címkehalmazból és *V* értékhalmazból építkező összes lehetséges dag halmaza rekurzíve így definiálható:

(i) $\varrho \in V \Rightarrow \langle \varrho, \{\varrho\}, \emptyset\rangle \in \Delta_{L,V}$
(Magyarázat: minden olyan gráf, amelynek egyetlen csomópontja van és nincs éle, egyúttal dag.)

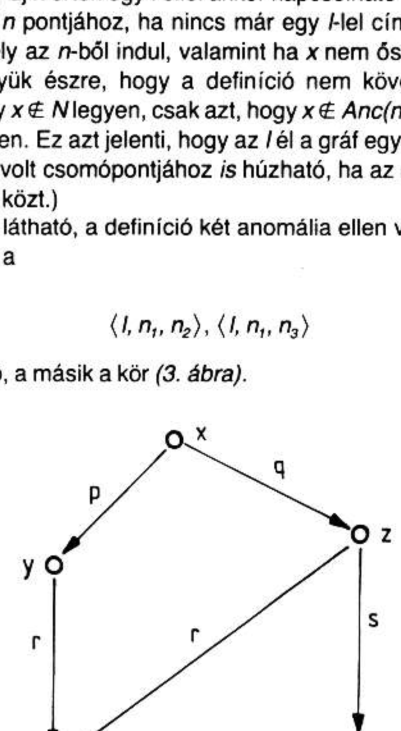
(ii) $\langle \varrho, N, E\rangle \in \Delta_{L,V} \;\&\; l \in L \;\&\; n \in N \;\&\; x \in V \;\&\; \nexists(n' \in N: \langle l, n, n'\rangle) \;\&\; x \notin Anc(n, \langle \varrho, N, E\rangle) \Rightarrow \langle \varrho, N \cup \{x\}, E \cup \{\langle l, n, x\rangle\}\rangle \in \Delta_{L,V}$

(Egy új *x* érték egy *l* éllel akkor kapcsolható a $\langle \varrho, N, E\rangle$ dag *n* pontjához, ha nincs már egy *l*-lel címkézett él, amely az *n*-ből indul, valamint ha *x* nem őse *n*-nek. – Vegyük észre, hogy a definíció nem követeli meg, hogy $x \notin N$ legyen, csak azt, hogy $x \notin Anc(n, \langle \varrho, N, E\rangle)$ legyen. Ez azt jelenti, hogy az *l* él a gráf egy korábban megvolt csomópontjához *is* húzható, ha az nincs az *n* ősei közt.)

Amint látható, a definíció két anomália ellen védekezik, az egyik a

$$\langle l, n_1, n_2\rangle, \langle l, n_1, n_3\rangle$$

szituáció, a másik a kör *(3. ábra)*.



3. ábra **Példa dagra**

A dagtávolság

Többféle lehetőség kínálkozik arra, hogy a tárgyszavakból épített dagok között, mint tetszőleges halmazelemek között távolságfogalmat vezessünk be. Illusztráció gyanánt hadd említsük meg az úgynevezett *diszkrét metrikát,* mely bármely halmaz fölött definiálható, éspedig:

$$\delta(x, y) = \begin{cases} 0, & \text{ha } x = y \\ 1, & \text{ha } x \neq y \end{cases}$$

Azt, hogy a fenti függvény metrika, az olvasó könnyen beláthatja. A mi javaslatunk egy a diszkrét metrikánál finomabb távolságfogalom lenne, de álljunk meg itt egy gondolat erejéig.

A továbbiakban a következő dilemma előtt állunk: Ha a matematika tisztaságát részesítjük előnyben a valósághű ábrázolással szemben, akkor nagyon szép struktúrát kapunk jól kezelhető matematikai tulajdonságokkal és könnyen bizonyítható tételekkel, de olyan szublimáltan tiszta távolságfogalommal, mely a valóságot nem igazán hűen tükrözi vissza. Egy példa: legyen két tárgyszavunk, mondjuk a *kőolajfinomítás* és a *számítástechnika.* Legyen továbbá egy színezett él, mely a *számítástechnika használata a kőolajfinomításban* viszonyt fejezi ki. Ez tehát egy egyszerű dag, mely születhetett például egy könyv osztályozásakor. Amikor a matematikai modellalkotás teljesen absztrakt, akkor a *kőolajfinomítás használata a számítástechnikában* dag ugyanolyan jogosultsággal jön létre, mint az előző, és nincs a rendszerben semmi olyan megkülönböztetési lehetőség, mellyel az előzőt preferálhatnánk az utóbbival szemben. Ez a tény hamis ekvivalenciákat és olyan bejárásokat visz be a dagok kapcsolatát leíró rendszerbe, amelyek később mint információs zaj köszönnek majd vissza a keresőrendszer használatakor. Másfelől, ha bármilyen szemantikát fogunk belevinni a tárgyszavazásba, veszíteni fogunk a matematikai következtetések terén – a gyakorlati nehézségekről most nem is szólva. (Gondoljunk csak arra, hogy ez a megszorítás azt jelentené, hogy minden tárgyszót felruháztunk különböző színű hurkokkal és kampókkal, hogy aztán csak a megfelelőkkel kapcsolódhassanak egymásba. Elegendő, ha csak a hihetetlenül megnövekvő tárigényre gondolunk.)

Bevalljuk, az absztrakt modell bemutatására szorítkozunk, mégpedig három okból. A dolgozat célja egy keret megadása a későbbi továbbgondolkodás elősegítésére; nem titkoljuk: hivatkozási alappá szeretnénk válni. A második ok prózaibb. Nincsenek meg egyelőre azok a konkrét statisztikai eredmények, amelyekkel egy „hurkos, kampós" rendszer bármiféle kialakítását alá lehetne támasztani. Végül a harmadik ok a legprózaibb. Szemantika bevezetésével a matematikai leírás embertelenül megnehezül.

(2) alatt már bemutattunk egy relációfajtát, az ekvivalenciarelációt. Most egy újabb relációtípussal ismerkedünk meg. Azt mondjuk, hogy $R \subseteq H \times H$ *parciális rendezés,* ha

(R) $\forall x \in H: (x, x) \in R$ *(reflexivitás)*

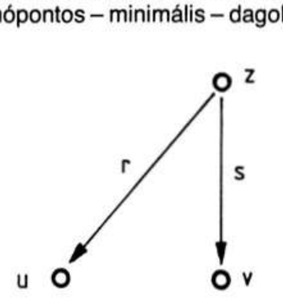
(A) $\forall x, y \in H: ((x, y) \in R \,\&\, (y, x) \in R) \Rightarrow y = x$ (6)
(antiszimmetria)

(T) $\forall x, y, z \in H: ((x, z) \in R \,\&\, (z, y) \in R) \Rightarrow (x, y) \in R$
(tranzitivitás)

Az a kifejezés is használatos, hogy ilyenkor *H részben rendezett halmaz.* Ha R parciális rendezés, indokolt az $(x, y) \in R$ helyett az $x \leqslant y$ használata.

Ismertetünk néhány szóhasználatot. Azt mondjuk, hogy egy részben rendezett halmazban x és y elemek *összemérhetőek,* ha vagy $x \leqslant y$, vagy $y \leqslant x$ fönnáll. Ellenkező esetben *összemérhetetlenek.* Azt mondjuk továbbá, hogy egy H részben rendezett halmazban *m minimális* elem, ha $\nexists n \neq m \in H: n \leqslant m$. Magyarul, ha nincs olyan, amelyik nála kisebb. (Felhívjuk a figyelmet a minimális és a *legkisebb* közötti különbségre: minimális az, amelyiknél nincsen kisebb, legkisebb az, amelyik mindenkinél kisebb. Legkisebb elemből csak egy lehet, minimálisból lehet több is. A legkisebb egyúttal minimális is, de a minimális általában nem legkisebb.) Egy részben rendezett halmaz minimális elemeit szokás *atomoknak* nevezni. Ugyancsak széles körben használatos a következő elnevezés: H részben rendezett halbazban b elemet *a rákövetkezőjének* nevezünk, ha $a \leqslant b \,\&\, a \neq b \,\&\, \nexists\, c \in H: c \neq a, c \neq b, a \leqslant c \leqslant b$. Vagyis, ha b úgy nagyobb, mint a, hogy „nincs közöttük senki".

Ennyi bevezető után módunk van a dagok között definiálandó parciális rendezésről beszélni. Maga a reláció nagyon egyszerű lesz, mégpedig a *részdagság.* Legyen $D_1 = \langle \varrho_1, N_1, E_1 \rangle$ és $D_2 = \langle \varrho_2, N_2, E_2 \rangle$ két dag; akkor mondjuk, hogy D_1 részdag D_2-ben ($D_1 \leqslant D_2$), ha $N_1 \subseteq N_2$ és $E_1 \subseteq E_2$. A részdag tehát olyan részgráf, mely maga is dag. Ennek a parciális rendezésnek az atomjai a tárgyszavak, mint egycsomópontos – minimális – dagok *(4. ábra).*



4. ábra **A 3. ábrán szereplő dag egy részdagja**

[4] nyomán abszolút atomosnak nevezzük a fenti rendszert, ha benne $D = \langle \varrho, N, E \rangle$ egy dag, $A = \langle a, \{a\}, \emptyset \rangle$ egy $a \notin N$ tulajdonságú atom, és ekkor $\langle \varrho, N \cup \{a\}, E \cup \{l, n, a\} \rangle$ D rákövetkezőjeként beemelhető a dagok halmazába valamely $l \in L$ él segítségével. Ez a fogalom azt fejezi ki, hogy egy tetszőleges daghoz egy benne nem szereplő csúcs valamilyen alkalmas éllel mindig hozzávehető, és az így keletkezett dag közvetlenül az előbbi fölött áll részgazdagság tekintetében.

Az abszolút atomosságról annyit mondhatunk, hogy tiszta matematikai eszközökkel vizsgálva nyilván teljesül, gyakorlatilag viszont valószínűleg nem. Itt jelentkezik a korábban emlegetett dilemma.

Nos, ha fenntartásainkat hangsúlyozva maradunk a dagok abszolút atomos részben rendezett rendszerének föltételezésénél, akkor a dagok közötti távolságra a nagyon egyszerű, *0–1–2 metrika* adódik. E metrikára:

$$\delta(D_1, D_2) = \begin{cases} 0, & \text{ha } D_1 = D_2 \\ 1, & \text{ha } D_1 \neq D_2, \text{de } \textit{összemérhetőek} \\ 2, & \text{ha } \textit{összemérhetetlenek} \end{cases}$$

Érdekesség, hogy annak bizonyításához, hogy δ valóban metrika, nem is kell tudnunk, hogy micsodák az elemek, amelyeken értelmezték. A bizonyítás az értékkészletből következik, hiszen nem lehet „kirakni" 1, illetve 2 hosszú élekből olyan háromszöget, amely a háromszög-egyenlőtlenséget megsértené.

Az „összeterelés"

Miután megvan a dagok közötti távolság, együttesen kell érvényesíteni a csomópontok közötti kétféle távolsággal:

Két dag a benne szereplő csomópontokkal együtt mint két vektor szemléltethető, melyek között, mint létrafokok sorakoznak az egymásnak megfeleltethető mennyiségek távolságai. Ekképp egy távolságvektor alakul ki. Ennek a távolságvektornak egyik komponense a dagtávolság, a többi az atomok (tárgyszavak) távolságaiból adódik. Ha egy ilyen távolságvektorral normát alkalmazunk, (valós) számot kapunk, mely ilyenformán *a két dag könyvtári értelemben vett távolságát* méri. Hogy ez valóban metrika, könnyen látható, hiszen, ha a két dag teljesen azonos, akkor a távolságvektor nullvektor, nullvektor normája pedig nulla; a szimmetria kézenfekvő a metrikák és a norma szimmetriatulajdonságai miatt; végül a háromszög-egyenlőtlenséget egy n távolságot tartalmazó távolságvektorral mutatjuk meg:

$$\begin{aligned} &\|\delta_1(x_1, z_1), \dots, \delta_n(x_n, z_n)\| + \|\delta_1(z_1, y_1), \dots, \delta_n(z_n, y_n)\| \geq \\ &\geq \|\delta_1(x_1, z_1) + \delta_1(z_1, y_1), \dots, \delta_n(x_n, z_n) + \delta_n(z_n, y_n)\| \geq \qquad (7) \\ &\geq \|\delta_1(x_1, y_1), \dots, \delta_n(x_n, z_n)\| \end{aligned}$$

Elértük végcélunkat: definiáltunk egy távolságot a könyvtár tételeit leíró klasszifikációs dagok halmazán, melyek közül egy (vagy esetleg több) ír le egy bibliográfiai tételt, s melyek közül egyet a könyvtárhasználó a keresés során megszerkeszt. Ezek után lehetőség van arra, hogy a kérdésdag valamekkora sugarú környezetéről beszéljünk, és kérhetjük egy számítógépes implementációtól, hogy az ebbe a környezetbe eső találatokat bocsássa rendelkezésünkre.

Összefoglalás

A rendszer összeállt, tekintsük át, hogy hol lehet „belenyúlni". Hat helyet fogunk fölsorolni. Ezek közül három olyan, hogy a matematikai apparátus változatlanul hagyásával a könyvtáros változtathat a tulajdonképpeni adathalmazon, illetve maga az élet. Három pedig a matematikai apparátust érinti:

Az előbbiek:

a) A rendszer önmagát hangolja, mert a tárgyszavak, mint valószínűségi változók, valamint kovarianciáik a könyvtár állományának változásával maguk is változnak;

b) a könyvtáros belenyúlhat a tezauruszba;

c) a (7) alatt bevezetett norma egy súlyvektorral elhúzható akár a dagtávolság, akár atomi építőkövek között mért távolság irányába;

illetve az utóbbiak:

d) választhatunk a kovariancia helyett alkalmasabb skaláris szorzatot a valószínűségi változók felett;

e) értelmezhetünk más metrikát a dagok felett;

f) s végül a (7) alatt bevezetett függvény nem kell, hogy szükségképpen norma legyen.

Zárszó

„Ott, ott a part!" – mutatja az Öreg Tengerész. Igen, munka, az lesz bőven. Nem szóltunk most – csak például – a számítógépes implementáció kérdéséről. Elgondolható, hogy e sorok írója is öreg tengerész lesz, mire a magyar könyvtáros-társadalom flottája kiköt e partokon. Néhány cirkálót mégis érdemes már ma előreküldeni.

Irodalom

[1] KÁLMÁN László: Dags and Geometries, Research Institute of Linguistics, MTA, Computational Linguistic Project, Document No. 10., Budapest, March 1990.

[2] CSABAY Károly: An Idea on Feature Geometries, Research Institute of Linguistics, MTA, Computational Linguistic Project, Document No. 11., Budapest, April 1990.

[3] KARVASZ Gyula: Analízis III. Jegyzet. Tankönyvkiadó, Budapest, 1978.

[4] FÁY Gyula–TŐRÖS Róbert: Kvantumlogika. Gondolat, Budapest, 1978.

Beérkezett: 1992. IX. 14-én.