a könyvtáros meghatározza igényeit és céljait, megkísérli meghatározni azt a továbbképzési típust, amely kielégíti;

a szekciókban működjenek olyan információs specialisták, akik a könyvtáros számára *felvilágosítást adnak a hozzáférhető programokról.* Ehhez adatbázist tartanak fenn a továbbképzési lehetőségekről és szoros kapcsolatban állnak más érdekelt szervekkel;

az *SLA Képzési Bizottsága* tudomással bír valamennyi szekció tevékenységéről és *a szekción túli lehetőségekről informál.* Más szakmai szervezetek programjairól is gyűjti az adatokat. Vizsgálja a könyvtárosok igényeit és ezek alapján döntéseket hoz. Továbbképzési programokat készít elő, köztük széles körben terjeszthető videoszalagokat, levelező tanfolyamokat, mozgó oktatási intézményeket, zártláncú televíziós programokat stb.

A bizottság támogathatja a továbbképzéssel foglalkozó szakfolyóirat kiadását, amely a képzési események hírei mellett szakcikkeket is közöl.

A javaslat végrehajtásához szükséges szervezet lényegében már létezik. A képzési programok koordinálásának lehetőségeit más szervezetek már sikerrel kipróbálták. Ezek szerint *a javasolt modell megvalósítható.*

Eredmények

A továbbképzés programjából néhány hosszútávon érvényesülő eredmény származhat:

- *az egyén számára:* önképzés, főként az elmaradás megakadályozása; a felhasználók jobb kiszolgálása;
- *a szervezet számára:* jobb munkateljesítmény; intellektuálisan megfelelő környezet a dolgozók számára;
- *az SLA számára:* a továbbképzés kívánatossá és hozzáférhetővé tétele, ennek alapján elérhető, hogy a könyvtáros ne találhasson kibúvót a továbbképzésből való kimaradásához;
- *a szakma számára:* a könyvtáros szakma a többihez hasonló biztonsággal alapozza meg önmagát, magasabb igényekkel és állandó emelkedéssel; javul a szakma presztizse; az egyesületek jobban válogatják és értékelik ki programjaikat; a munkaadók támogatják és megbecsülik a könyvtárosok továbbképzési igényeit.

/KIRK, A. G.: A model for continuing education for special libraries = Special Libraries, 67. köt. 3. sz. 1976. p. 138–144./

(Sárdy Péter)

OSZTÁLYOZÁS – INDEXELÉS

A gépi indexelés nyelvelméleti és szemiotikai problémái

A számítógépek egyre hétköznapibbá válása kapcsán felmerül a kérdés, vajon *képesek-e a gépek a gondolkodás munkájának átvételére az embertől,* vagy ahogy A. M. TURING megfogalmazta: működhetnek-e oly módon a gépek, mint ahogy mi viselkedünk, amikor azt mondjuk, hogy gondolkozunk.

Más szavakkal kifejezve e gondolatot: *eléggé ismerjük-e mi az emberi gondolkodás folyamatát ahhoz, hogy ezt explicit formában előírhassuk egy gép számára?*

Az elmúlt évtizedben az információtudomány területén sokat foglalkoztak azzal a problémával, miként lehetne a számítógépeket természetes nyelvű szövegek feldolgozására felhasználni. A számítógépek ilyen alkalmazása igen jelentős az információkeresés szempontjából. Ehhez olyan módszerek kifejlesztése szükséges, amelyek *a számítógépet alkalmassá teszik az indexelés, a referátumkészítés és a keresés szellemi feladatainak megoldására.* Mivel a kommunikáció legfontosabb eszköze az emberi természetes nyelv, ezért a kutatás egyik fő irányát jelenti a természetes nyelvű kommunikáció és a digitális számítógépek kompatibilitásának kidolgozása.

Mivel a gépi indexelés, referátumkészítés és szövegkeresés a mondanivalónak a szövegből való kivonatolását feltételezi, e folyamatok gépesítése annak szükségességét fejezi ki, hogy a számítógép szinte megértse a szöveg jelentését. Nyilvánvaló, tehát hogy *a jelentéssel, az értelemmel kapcsolatos vizsgálatok az automatikus szövegfeldolgozás-kutatás egyik legfontosabb részének tekinthetők.* Ehhez nem elegendők az empírikus közelítések, a további kutatómunkához alapos elméleti ismeretek szükségesek. A mélyebb ismeretek megszerzésének egyik fő iránya a nyelvelméleti és szemiotikai kutatás, mind az emberi, mind a gépi indexelés vonatkozásában.

Gépi indexelés

A gép a természetes nyelven írott szövegeket egymást követő *szimbólumokként* kezeli. Azonosítani tudja a szóközzel elválasztott karakterláncokat, igen/nem döntésekre képes a karakterek jelenléte, hiánya és egymáshoz viszonyított helye tekintetében. Mivel a gép emberi értelemben nem ismerheti fel a szavak jelentését, számára olyan kulcsokat kell kidolgoznunk, mint a szavak vagy szókészletek előfordulási gyakorisága egy szövegben, együttes előfordulásuk, viszonylagos helyzetük mondatokban, szavakat alkotó karakterláncok megkülönböztető jellege stb.

Az előfordulási gyakoriság szerinti értékelés alapja, hogy ez feltétlenül jellemzi a szöveg tartalmát. A

számítógépet úgy programozzák, hogy gyűjtse ki a jellemző szavakat, előfordulási gyakoriságuk csökkenő sorrendjében. Ebből a jegyzékből választják ki az indexszavakat, kifejezéseket. Ezt a módszert még kombinálják a szöveg jelentősebb részeinek (a legtöbb hasznos információt tartalmazó szövegtestek) kijelölésével is.

Egy másik gépi indexelési eljárásban a szöveg szavait és mondatait olyan szavak és mondatok jegyzékével hasonlítják össze, amelyeket a dokumentum *speciális indexkifejezéseinek* generálására szánnak.

A következő jelentős gépi indexelési eljárás lényege, hogy a géppel azok a szavak ismerhetők fel, amelyek *láncképe* megegyezik az indexelés számára érdemi információkat hordozó szavak láncképével. Ez lehet a legjellemzőbb szakkifejezések karakterisztikus töredéke is.

A gépi indexelés kutatása a szemiotikának főleg a szintaktikai összetevőjére koncentrált, vagyis *a jelek egymással való kapcsolatainak vizsgálatára.* Kisebb mértékben foglalkoztak a szemiotika szemantikai szintjével, vagyis *a jelek és az általuk jelölt fogalmak* – az értelem, a jelentés – *kapcsolatával.* Olyan algoritmusokat eddig még nem sikerült kialakítani, amelyek – az emberhez hasonló módon – a gép számára lehetővé teszik a jelentés megállapítását is, a jelek és jelsorozatok felismerése és elhelyezése útján; ez tisztán szellemi feladat.

Indexelés mint kommunikációs folyamat

Az indexelés művelete során dokumentumok halmazát állítják elő, a halmaz minden tagjára érvényes tartalmi leíró fogalmakkal. Az *indexkifejezésekkel helyettesítjük a dokumentumok teljes szövegét.* Az indexelés magától értetődő velejárója az *információveszteség,* ami csökkenthető vagy optimálható, de teljesen nem szüntethető meg. Az indexelés szükségessé teszi a szöveg jelentésének megértését és az információtartalom értéke megítélésének képességét, az információ várható keresőinek, használóinak szempontjából.

Az emberi indexelés műveletét főleg *a szemiotika szemantikai szintjén vizsgálták.* Az indexelés azonban ezen túlmenően kommunikációs folyamatot jelent. A pragmatika szintjén az információs folyamatok olyan kölcsönhatásokkal jellemezhetők, amelyek az átvitt információk, ezek felhasználói és tolmácsolói, és az információ végső célja (amelynek szolgálatára hivatott) között jönnek létre. A szöveg jelentése ezen a szinten a realitások, a mindennapi élet, az emberek közötti kommunikáció kapcsán *valakinek* bizonyos alkalomból szóló jelentésnek tekintendő.

A jelentés, az értelmet hordozó nyelv egy kommunikációs kódja, noha nem tökéletes kód. Ez viszont kevés a gondolatok átviteléhez. Az írott és beszélt nyelv jelentését nemcsak a nyelvtani rend, valamint a szavak és tárgyak, ezek tulajdonságai stb. közötti összefüggések adják, hanem függ a körülöttünk levő világ szerkezetétől is, megszerzett tapasztalatainktól is. Tehát ilyen szinten az indexelő munkájához tartozik egy sereg nem nyelvelméleti tényező figyelembe vétele, mint a szerzők célja, motivációi, háttere stb; mindez a dolog *kommunikációs* oldalát alkotja.

Redundancia a természetes nyelvekben

A nyelv komplex szintaktikai szabályai olyan megszorításokat hoznak létre, amelyek redundanciára vezethetnek. A szintaktikai redundancia a szöveg kiegészítését jelenti: kissé többet mondunk vagy írunk, mint amennyi feltétlenül szükséges lenne egy üzenet közvetítéséhez. *A szemantikai redundancia a nyelv elégtelenségéből származik,* ami azzal az igénnyel jár, hogy kifejezéseinket, mondatainkat bővítsük, gondolatainkat különféle módokon fejezzük ki. Ezzel együtt jár az ismétlés vagy felesleges jelek használata.

A redundanciának fontos következményei lehetnek a gépi indexelésben. Egyes gépi módszerek a szintaktikai és szemantikai redundanciát úgy csökkentik, hogy *bizonyos szavakat vagy meghatározott hosszúságú szavakat elhagynak az indexelendő szövegből.* A szemantikai redundancia azonban hasznos is lehet, mert egy biztonsági tényezőt hoz magával. Ezért ez felhasználható az indexelési algoritmus optimális kialakítására.

A nyelvelméleti kutatások és a gépi indexelés

Az információtudomány elismert ténye, hogy az automatikus szövegfeldolgozáshoz általában is, de különösen a gépi indexeléshez, a fejlődésükhöz szükséges ismeretanyag nagyrészét a nyelvelméleti kutatásoknak, vagyis *a nyelvek szerkezetével és működésével foglalkozó vizsgálatok* eredményeinek kell szolgáltatnia. Erre vonatkozólag a kutatások különféle irányzatairól beszélnek, köztük a *pszicholingvisztikáról,* amely az üzenetek és ezek válogatását, értelmezését végző egyének közötti kapcsolatokkal, vagyis a nyelvészeti jelek és felhasználóik közötti kapcsolatokkal foglalkozik. E kapcsolatokat tekintjük az információs folyamatok központi problémájának a szemiotika pragmatikai szintjén is.

Az üzenetben a szerző szándéka a kulturálisan *elfogadott kód szerinti jelekké alakul át.* Az indexelés az üzenet jelentésének megértését és ennek új üzenettel (az index-kifejezésekkel) való helyettesítését foglalja magában. Ebben a folyamatban az indexelő – akár gép, akár ember – képviseli mind a forrást, mind a címzettet, vagyis az üzenet kódolóját és dekódolóját egyaránt.

A gépi indexelésre érvényes algoritmusok kifejlesztése tulajdonképpen a szöveget kitevő jelek és a szöveg jelentése közötti összefüggések megértését jelenti. Azonban itt nem elegendő az egyes szavakhoz azok jelentésé-

nek hozzárendelése és a jelentés összetevőinek – nyelvtani alapokon végzett – mondatokká kombinálása. A szöveg értelmének megértése túlmegy a szöveg által nyújtott nyelvészeti jelentésen, mivel *az nem nyelvészeti, hanem pszichológiai kategória.*

A nyelvelméleti kutatások erősen szintaktikai orientáltságuk miatt nem tettek jelentős előrelépést az emberi nyelvben a jelentés természete és lényege megértésének központi problémája megoldásában. A nyelv tágabb szemléletére van szükség. A természetes nyelvű szövegek információfeldolgozási célokat szolgáló gépi kezelése kapcsán felmerülő legfontosabb elképzelések arra utalnak, hogy *a pszicholingvisztika és szemiotika közös területei által nyújtott természetes nyelv vezethet el olyan alapvető ismeretekhez, amilyeneket az automatikus szövegfeldolgozás kutatói igényelnek és keresnek.*

/ARTANDI, S.: Machine indexing: linguistic and semiotic implications = Journal of the American Society for Information Science, 27. köt. 4. sz. 1976. p. 235–239./

(Roboz Péter)

DEWEY, az osztályozási rendszerek és a szemantikai univerzum

A haladás és az interdiszciplináris kutatás között szoros a kapcsolat, mivel az egyes diszciplinák közötti üres területek analógiákkal, hasonlóságokkal és szintézisekkel kitölthetők. E területen *nagyon jelentős M. DEWEY tevékenysége, aki 100 éve alkotta meg a Tizedes Osztályozás rendszerének alapjait.* Ezt a rendszert fejlesztette tovább – mintegy 20 évvel később – OTLET ÉS LA FONTAINE. A továbbiakban erről a továbbfejlesztett rendszerről, azaz az ETO-ról, az *Egyetemes Tizedes Osztályozásról* lesz szó.

Az ETO azonban nem terjedt el kellő mértékben; ennek egyik oka az, hogy *nem ismerik elegen*, a másik pedig az, hogy százéves léte során nem eléggé szisztematikusan és nem a megfelelő mértékben fejlődött, és ezért *sokan hibásnak tartják.* E hiányosságokon kell segíteni szemantikai tartományokra osztással, ezek tömörségének meghatározásával és ETO-tezauruszoknak megfelelő alfabetikus jegyzékekkel való kiegészítéssel.

A szemantikai rendszer definiálásánál a résztvevő személyekre feltétlenül tekintettel kell lenni. A szemantikai tér matematikai értelemben annyi dimenzióból áll, ahány szemantikai egység alkotja a rendszert: mindegyik szemantikai egység egy n-dimenziós térben lévő soklapú test egy-egy csúcspontja; tehát nagyon sok szemantikai kapcsolat lehetséges. *Mindegyik bibliográfiai egységnek saját szemantikai tere van, amely meghatározott számú, egymással kapcsolatban lévő szemantikai egységből áll.* A térben az egyes bibliográfiai egységek és az egyes keresések mindenkor központi helyzetet foglalnak el; az összes többi egység és keresés ehhez viszonyítva másodrangú helyzetű. A különféle helyzetek között szemantikai távolságokból álló háló alakul ki.

A fogalomalkotás k-dimenziós térben zajlik le. Az átlagos emberi agyvelő k-értéke nem túl nagy, de egynél sokkal nagyobb. A k-érték okozza az egyes ismeret-tartományok közötti viszkozitást; és ezért tűnik úgy, hogy az osztályok képzéséhez szükséges különféle szempontok az ismeretek és a fogalmak stabil alakzataként rendelkezésre állnak. Az – a 17. század tudósainak reményeihez és illúzióihoz hasonló – elképzelés, hogy a teljes ismeretanyag rendszerezett összességének valóban birtokában vagyunk, általában akkor merül fel, ha nem ismerjük fel azt, hogy *a gondolatok és a valóság, illetve a gondolatok és a nyelv közötti kapcsolatok nem mindig egységesek és lineárisak.*

Az ETO-t a szemantikai térben 1 és 10 közötti osztásokkal ellátott egyenes ábrázolja. A fogalmakat *egyetlen szemantikai dimenzióban,* nem-periodikus, véges decimális számok sorozatában foglalják össze. Az ETO szemantikai érvényességét kérdésessé tevő probléma az, hogy milyen áron lehet a szemantikai dimenziók számát k-ról egyre csökkenteni; kielégítő módon ábrázolja-e az ETO a szemantikai univerzum összetettségét? Ezeket a kérdéseket most nem válaszolhatjuk meg, de az ETO rendszerében használt decimális számok tulajdonságairól még szólni kell.

Az ETO számfogalmai azonos számú számjegy esetén nem azonos szemantikai kiterjedésűek, mert a diszciplinák kezdeti felosztása eltérő. Ebből következik, hogy *ma félmilliónál kevesebb ETO-szám érvényes, bár 15 számjegyet használva, számuk lényegesen nagyobb lehetne.* Ennek következménye az, hogy számítógépes rendszerekben az ETO hatásfoka elenyészően kicsi, mintegy 10^{-9} nagyságrendű. Az ETO-osztályok betöltöttségének számítására vezették be a DUD fogalmat *(DUD – Densita di Utilizzazione Decimale = a decimális betöltöttség).* A DUD az ETO számok mennyisége és a megfelelő ETO számtartomány maximális számtartományának megfelelő számjegyszámú, természetes számok mennyisége közötti arányt adja meg.

A DUD(54) az 54 jelű ETO osztály DUD értékét jelenti, és egyszerűen kiszámítható. A $DUD_6(54)$ az 54 jelű ETO osztály hat számjegyből álló számaira terjed ki, és így az 540 000 és 540 999 közötti, kémiai fogalmak tartoznak hozzá. A $DUD_6(54)$ értéke 12%. A pszichológia tárgykör DUD_6 értéke pedig mindössze 2%. Így merül fel az a kérdés, hogy *megnövelhető-e a teljes ETO DUD értéke 80–100%-ra?* Ez nemcsak lehetséges, hanem az ETO rendszer gazdaságosabb számítógépes alkalmazásához szükséges is.

Az ETO számait rögzítő szemantikai egyenes pontjaihoz való ragaszkodás megakadályozza az ETO rendszer egyenletes fejlődését. *Olyan rendszert kell teremteni,*