INFORMÁCIÓKERESÉS

Rédey Gábor, *redeyg@iif.hu*
Országos Atomenergia Hivatal

Neumann Attila, *neumann.attila@chello.hu*
Neumann Fivérek Kft.

Sütő Zoltán, *suto.zoltan.tzm@freemail.hu*
Duna Info Kft.

Kivonat

A cikk óvatos kezdeményezés az információkeresés nyelvének és ezen keresztül az egész információkeresési folyamat „átértelmezésére".

Bár az emberi gondolkodás jórészt nyelvfüggetlen, ezzel szemben az emberi információcsere erősen nyelvhez kötött, ezért nyilvánvaló, hogy az információkeresés folyamata is nyelvfüggő. A természetes nyelvet önmagában információkereső nyelvként felhasználni ma még nem szokványos, hiszen ez az ötlet bár ígéretes, meglehetősen sok és nehéz problémát vet fel. Ennek ellenére az irodalomban kb. a 90-es évek eleje óta olvashatunk ilyen célú kutatásokról és eredményekről.

A hagyományos információkereső nyelvek a természetes nyelvekhez képest szegényes szintaktikai szerkezetűek. Az ezeken a nyelveken feltett, olykor „homályos" kérdésre sokszor „zajos" (irreleváns) vagy nagy információveszteségű válasz érkezik. Ebben a helyzetben előrelépést csak egy rugalmasabb szintaxisú információkereső nyelvtől várhatunk, amely nemcsak az egyedi fogalmakat (vagy azok valamilyen együttesét), hanem azok természetes nyelvi relációit is modellezni képes. A cikk célja, hogy betekintést nyújtson egy ilyen új típusú, finomabb felbontású információkereső nyelv részleteibe.

A cikk megmutatja, hogy léteznek a természetes nyelvek logikai finomstruktúráját hűen leképező ismeretreprezentációs nyelvek, amelyekre mind a természetes nyelvű szövegek, mind a háttérismeretek (ontológiák, tezauruszok) könnyen (majdnem automatikusan) lefordíthatóak, továbbá a keresőkérdés is ugyanezen a nyelven reprezentálható. A kérdésre generált válasz magyarázható: lépésenként nyomonkövethető.

1. Bevezetés

Hagyományosan, az információkeresés színhelye és tudománya a könyvtárak, a könyvtárosok, a könyvtártudomány ügye volt. A könyvtári információkeresés módszerei még ma is erősen kötődnek az információtárolás hagyományos technológiájához. A számítástechnika megjelenésével azonban ez a kizárólagosság fellazult. A számítástechnika eltömegesedése előtti korszakban a fejlődés még szerves volt (pl. Kunszt: LOGEL rendszere) [?]. Később, az új technológia rohamos elterjedése nyomán a háttérben újraértelmeződött egyebek mellett az információ és az információkeresés. Ez a helyzet mára némileg megváltozott (szemantikus web), bizonyos közeledés tapasztalható a két szemléletmód között.

2. Mi az információ?

Az információ olyan alapfogalom, amely több nézőpontból is tekinthető. Az egyik ilyen nézőpontot az információelmélet képviseli, amely a statisztikai valószínűség elvein alapul, és a kibernetika egyik ágának számít. Az információelmélet tárgya az információ mennyiségi vonatkozása, amely a mi szempontunkból kevéssé érdekes. Ehelyett érdeklődésünk középpontjában az információ minőségi oldala áll. Ezt a megközelítést, az előzővel ellentétben nevezhetjük szemantikai információnak is, melyet a következőképpen definiálnak az irodalomban.

> ... az információ valamely szövegnek olyan struktúrája, amely alkalmas arra, hogy változást idézzen elő a befogadó képstruktúrájában. [?]

> Információ fogalmán a viselkedést befolyásoló, új ismeretet nyújtó adatok tartalmi jelentését értjük. Az adatok és hírek csupán információhordozók. Az információ határozatlanságmennyiség-megváltozást okoz, jelentése az ismeretszint különbség. [?]

Az információ elvi meghatározásán túl nem vonatkoztathatunk el annak gyakorlati megközelítésétől sem, vagyis attól, ahogyan az információra a mindennapok során sok esetben nem tudatosan gondolunk.

> Az információ megjelenési formája általában természetes nyelvű szöveg, amelyből csak megfelelő szövegértelmező, illetve -feldolgozó képességgel tud a tanuló ismereteket szerezni. [?]

> ... információnak nevezünk mindent, amit a rendelkezésünkre álló adatokból nyerünk. Az információ olyan tény, amelynek megismerésekor olyan tudásra teszünk szert, ami addig nem volt a birtokunkban. Az információ legkisebb egysége a bit. A számítástechnikában a programok is 1 bites információkból épülnek fel. [?]

Könnyű észrevenni, hogy az információ elvi meghatározásában nincs nagy különbség a különböző szakterületek között. A különbség inkább az információfogalom gyakorlati megközelítésekor bukkan felszínre. A gyakorlati megközelítés pedig legjobban az információ reprezentálásának mikéntjében ölt testet.

Az eddig említett két megközelítési mód – amit az egyszerűség kedvéért „könyvtári" és „számítástechnikai" megközelítésnek neveztünk – az információ fogalom gyakorlati értelmezésében tér el alapvetően egymástól, és természetesen mindkettő az elvi megközelítéstől is.

3. Információreprezentáció

3.1. Hagyományos információkereső nyelvek

Az elvi megközelítés az általános emberi információ reprezentációhoz kötődik, a szöveg (adat) struktúrájában látja annak megjelenését.

Ettől az állásponttól távolodik el a könyvtári információ fogalom azzal, hogy az információ reprezentálására absztrakt információkereső nyelvet alkalmaz. A könyvtári információkereső nyelv olyan mesterséges nyelv, melynek szavai nem feltétlenül a természetes nyelv szavai, és az e szavak által reprezentált fogalmak bizonyos meghatározott relációk által részben rendezettek. Világos, hogy az információkereső nyelvek kifejező ereje a természetes nyelvekéhez képest romlik, nem beszélve az egyéb mellékhatásokról.

Hasonló jelenség figyelhető meg a számítástechnikai információ fogalom esetében is. A számítástechnika információ reprezentációja formális, bitközpontú. A számítástechnika számára egy szöveg mindössze bitek sorozata. A reprezentációs veszteség mibenléte itt is azonnal szembetűnő.

Összegezve megállapítható, hogy mindkét eddig tárgyalt megközelítés a lehetőségek talajáról, bizonyos veszteséggel reprezentálja az információt, ami eleve meghatározza a keresés minőségét és eredményességét.

Ezt a képet némileg árnyalják elsősorban a könyvtári információkereső nyelvek fejlesztési törekvései. A már említett Kunszt-féle kutatás [**?**] pl. célul tűzte ki a megszokott generikus, egész-rész stb. kétargumentumú relációk kiegészítését többargumentumú relációkkal, és ezzel megközelíteni a természetes nyelvek kifejezőképességét. Kívánatos lenne, hogy az információkereső nyelvek is képesek legyenek ezeknek az adott esetben érvényes szintaktikai összefüggéseknek a kifejezésére, vagyis erre alkalmas eszközökkel rendelkezzenek. A nem gépi információkeresés céljaira előállított eszközökben (például különféle speciális mutatókban) voltak és vannak erre szolgáló eszközök, de a gyakorlatban alkalmazott információkereső rendszerekben – legyenek azok akár hagyományos katalógusok vagy mutatók, akár online számítógépes információkereső rendszerek – ilyenek használata csak igen ritkán, kivételesen fordul elő [**?**].

A számítástechnikai megközelítés információ fogalma sem maradt változat-

lan. Az utóbbi évek során nagy mértékű közeledést tapasztalhatunk a könyvtári információ fogalomhoz, gondolunk itt különösen arra a változásra, ami a „teljes szöveges keresés”-től a „szemantikus web” fogalmáig vezetett. Ugyanakkor a két módszer a gyakorlatban meglehetősen el is különül egymástól, kialakult alkalmazási területeik inkább kiegészítik, mint átfedik egymást. Ez természetes módon felveti azt a problémát, hogy az információkeresés mégiscsak egységes szemléletű, nem függhet attól, hogy éppen mit, miben, milyen céllal keresünk. A következő rövid szakaszban ezt az eredeti célt tarjuk szem előtt.

3.2. Tudásreprezentációs nyelvek

Az eddigiek alapján felvetődik a kérdés: vajon létezik-e olyan gyakorlati információ fogalom, ami az előzőeknél jobban megközelíti az információ elvi értelmezését. Abból indulhatunk ki, hogy az információ emberi reprezentálására alkalmasabb eszköz, mint a természetes nyelv nem létezik. Ez indokolja, hogy a természetes nyelveket modellező mesterséges intelligencia rendszereket tekintsük az információt leghívebben reprezentáló nyelveknek, amelyek képesek az információ legmélyebb szemantikai összefüggéseinek tükrözésére.

A természetes nyelvek szemantikai információ tartalmának reprezentációja nagyon régi keletű törekvés, a logika tárgya. A modern szimbolikus logika kezdetét a XIX. század végétől számítják. Ez természetesen nem jelenti azt, hogy az ókori vagy a középkori logika eredményei mellőzhetők lennének, éppen ellenkezőleg, valójában messzemenően azokra az eredményekre támaszkodunk. Mindenesetre azzal az igénnyel, hogy a logikai következtetések az aritmetika módjára kiszámíthatók legyenek, először Leibniz lépett föl, célját azonban – legalábbis részben, a matematika nyelvére korlátozva – csak Frege érte el két századdal később. Mindezekkel arra utalunk, hogy a logikai ismeretreprezentáció célkitűzései és eredményei felelnek meg leginkább a szemantikai információ olyan igényű reprezentálásának, ami lehetővé teszi hogy adott esetben egy információbázis számára feltett szemantikus kérdés egyáltalán kiértékelhető legyen.

A logikai ismeretreprezentáció természetesen a logika nyelvén valósul meg. Ez a nyelv ma sokak számára a szimbolikus logika XIX. és XX. század fordulóján kialakult nyelvét jelenti. Ezt a nyelvet Frege az aritmetika nyelvének mintájára alkotta meg. A természetes nyelvek és az aritmetika nyelve azonban bonyolultságban nagyon eltérőek. A hagyományos logika nyelve – bár voltak erre kísérletek – nem alkalmas a természetes nyelvek logikai szerkezetének modellezésére. Nem azért, mert a feladat nem volna így megoldható, hanem mert az eredmény gyakorlatilag nem használható [**?**].

Egyetlen fiú sem csak Marit szerette.
$P^{\rightarrow}\{\sim \exists x(\textit{fiú}\ x)\ \&\ [(\lambda y.\textit{szeret}\ x\ y)\ EQU\ \lambda y(y = Mari)]\}$

A fenti illusztráció azt mutatja, hogy a formula előállítása és visszaolvasása egy-

aránt nehézséget okoz. Az algoritmus, amelynek segítségével egy természetes nyelvű (típus)mondatból a hozzá tartozó logikai formula előállítható, kilátástalanul bonyolult.

Olyan logikai nyelvre van tehát szükség, amely nemcsak a természetes nyelvű mondatok szemantikai tartalmának hű leképezésére képes (mint ahogyan a hagyományos logika nyelve), hanem a nyelv szintaktikai viszonyainak hű leképzésére is. Ekkor ugyanis elvárható, hogy (alkalmas természetes nyelvi elemző közbeiktatásával) a természetes nyelvű mondat szintaktikai egységei könnyen átfordíthatók a logikai nyelv szintaktikai egységeire. Vagyis a *természetes nyelvű szöveg* $\leftrightarrow$ *reprezentált szöveg* közötti fordítás gépesíthető.

A vázolt problémára az irodalomban létezik több megoldás is. Anélkül, hogy a részletekbe bocsátkoznánk, csak egy-egy példát villantunk fel az egyes módszerek legszembetűnőbb sajátosságainak illusztrálására. A részletek iránt érdeklődők számára az irodalomra utalunk.

Sowa *conceptual graphs* (CG) néven ismert reprezentációs nyelvét [**?**] Peirce egzisztenciális gráfnyelvéből vezeti le:

All trailer trucks are eighteen wheelers.
`[trailerTruck : ∀] → (part) → [wheel : {*}@18];`

Iwańska *UNO* nyelve [**?**] kifejezetten a nyelvtani szerkezetre épít, bizonyos alaprelációkkal kiegészítve:

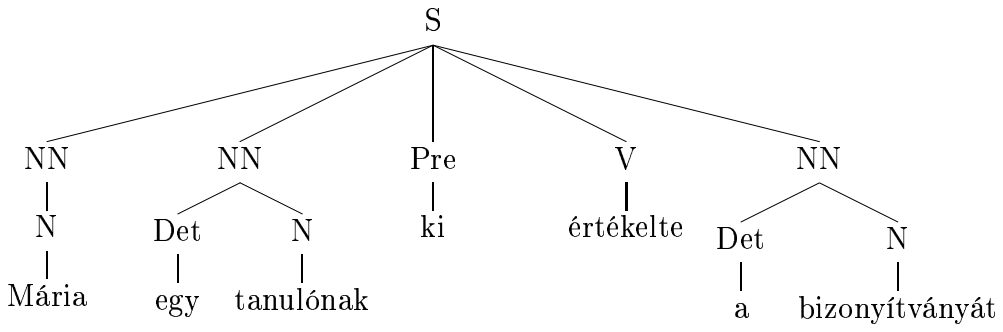
Every student works hard.
`np(det ⇒ every, n ⇒ student) == [work(adv ⇒ hard)];`

Rédey *intenzionális szövegreprezentációs* nyelve (iCTRL) [**?**] szintén a mondat nyelvtani relációit tükrözi, abból az alapfeltételezésből kiindulva, hogy a nyelvtani szerkezet a logikai szerkezetet teljes egészében magában foglalja:

Mária egy tanulónak kiértékelte a bizonyítványát.

```
((((értékelte x y)z w,
  a y, bizonyítványát y),
  ⟨ki⟩w),
  egy z, tanulónak z),
  Mária x.
```

Azt, hogy a szintaxis a természetes nyelvek alapvető nyelvtani relációira (állítmány, alany, tárgy, határozók) épül, példamondatunk nyelvtani elemzése szemlélteti:



1. ábra. A *Mária egy tanulónak kiértékelte a bizonyítványát* mondat nyelvtani elemzése

A nyelvtani és logikai szerkezet ilyen szoros kapcsolata garantálja, hogy a *természetes nyelvű szöveg* $\leftrightarrow$ *reprezentált szöveg* fordítás gépi úton valóban könnyen végrehajtható. A gépi reprezentációra fordítás lehetősége olyan mozzanat, amelynek hiánya értelmetlenné tenné a szóban forgó reprezentációs nyelv minden más esetleges előnyét. Emellett ugyanebből, tehát, hogy a reprezentáció mind szintaktikailag, mind szemantikailag, minden részletében követi a természetes nyelv szerkezetét, következik, hogy a reprezentált szöveg minden részlete elvileg a keresés számára hozzáférhető.

A következőkben vázoljuk az információkeresés elvét, továbbá a szöveghű ismeretreprezentációs nyelvekre alapozható információkereső rendszerek architektúráját.

4. Mi az információkeresés?

Információkeresésen általában azt értik, amikor valamilyen formalizált információt hasonlítanak a már meglévő, rendelkezésre álló formalizált információ halmaz elemeihez. Ennek hátterében az áll, hogy a keresést mindig valamilyen tudáshiány váltja ki, ami vagy valamilyen feltételezés formáját ölti, amelynek ismeretlen igazságértékét verifikálni kell, vagy valamely, bizonyos konkrét tulajdonságokkal rendelkező ismeretlen létezésének a feltételezése, amit szintén igazolni kell. A keresés maga mindig valamilyen előzetes, többnyire nyilvánvalónak gondolt (ezért általában nem kifejezett) ismeretre épül. A keresést követően a talált információ optimális esetben növeli a kereső már meglévő ismeretszintjét.

Az előzőek alapján nem meglepő tehát, hogy a kereső nyelv sajátosságai meghatározzák a keresés eredményének várható minőségét is. Ha csak karaktersorozatot tudunk keresni egy másik karaktersorozatban, akkor annál többet nem várhatunk, mint hogy meg is találjuk. Ha a kereső nyelvünk szavai részben rendezettek bizonyos relációkra nézve, akkor jogosan feltételezhetjük, hogy ez a keresés eredményében is tükröződik.

A keresés minőségét alapvetően befolyásolja az a háttérismeret, amire támaszkodni lehet. A puszta karaktersorozat keresés esetében semmilyen háttérismeretet nem tudunk felhasználni, ellenben a könyvtári információkeresés vagy a szemantikus web masszív háttérismeretre támaszkodik. Ez a háttérismeret azonban általános és statikusan rögzített, csak lassan bővül. Az az információtömeg, ami a keresés bázisát jelenti, nem, vagy csak viszonylag nagyon szűk hányadban vesz részt a keresésben. Mivel a keresés mindig csak a reprezentációs (információkereső) nyelven hajtható végre, ezért ez más megvilágításban azt jelenti, hogy a hagyományos információkeresés számára a szöveg jelentős része „láthatatlan" marad, pontosan annyi információ érhető el a keresés során, amennyit a reprezentációs nyelv „felbontó képessége" megenged. A reprezentációs nyelv tehát eleve meghatározza a keresés minőségét, ami magától értetődően alátámasztja az információkereső/-reprezentációs nyelv célszerű megválasztásának alapvető jelentőségét: „... az információ annyit ér, amennyi megtalálható belőle"[1].

5. Egy lehetséges kérdés–válasz rendszer

A következőkben vázolunk egy olyan rendszert, amely minőségi előrelépést jelent az információkeresésben. Az alkalmazott ismeretreprezentáció módszere tekintetében nincs korlátozás, elvileg bármely, szövegek szemantikai információtartalmának reprezentálására alkalmas módszer alkalmazható. Ilyenek pl. a fent már megemlített CG illetve az UNO reprezentációs módszerek, továbbá az iCTRL, amelyet munkánk során alkalmazunk.

Egy ilyen rendszerrel szemben a következő követelményeket állítjuk:

1. A rendszer képes megtalálni bármely természetes nyelven megfogalmazott terminust, és megmutatja, hogy azt mely szövegösszefüggésben találta.

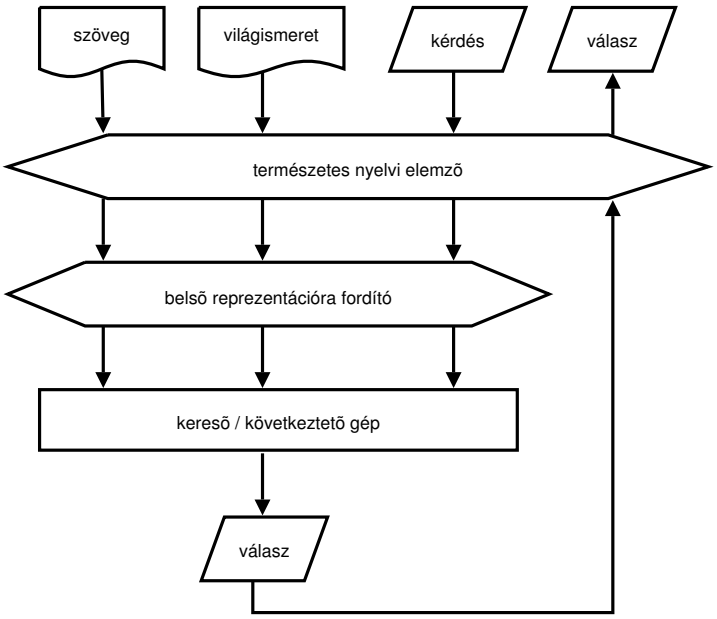
2. A terminus keresése közben lehetőség van asszociációkra (az eredetivel valamilyen relációban lévő terminusok keresésére).

3. A rendszer számára természetes nyelvű kérdéseket lehet megfogalmazni, és azokra ugyanazon a nyelven válasz érkezik.

4. A rendszer megmagyarázza, hogy az adott kérdésre adott válaszhoz milyen közbenső lépésekben jutott el.

5. A kérdések megfogalmazásához segítséget nyújt a rendszer: az ember számára emészthető formában mutatja meg az általa használt fogalmakat és a közöttük lévő összefüggéseket.

[1] Prokné Palik Mária: A tartalmi feltárás problémái online Könyvtári katalógusokban, BME OMIKK



2. ábra. A kérdés–válasz rendszer működési sémája

Mi egy ilyen rendszer lényege? Működtet egy „értelmező motort”, mely képes értelmezni egy természetes nyelven megfogalmazott állítást vagy kérdést, sőt képes értelmezi természetes nyelven tárolt szövegeket is, továbbá a kérdés valamint a szövegek között a kívánt relációk szerinti megfeleltetést képes létrehozni, vagyis szövegben szemantikus információt keresni. Lényegében ez azt jelenti, hogy egy gép, egy nagy információhalmazt olvas át az ember helyett, és képes abból kiemelni pl. a kívánt relációknak megfelelő szövegrészeket. Az információkeresésnek ez az elképzelhető legkényelmesebb módja. Az ember röviden elbeszélget egy géppel, és eredményként megkapja egy nagy szöveghalmazból a számára fontos szövegrészeket, a feltett kérdéseire adott válaszokat. Mindez, bár meglehetősen futurisztikusan hangzik, megoldható az előbbiekben vázolt nyelvi elemző- és reprezentációs módszerek alkalmazásával, amelyek jól követik azt az absztrakciós folyamatot, amit az ember a természetes nyelvekben használ.

6. Az információkereső rendszerek hatékonysága

A mesterséges intelligencia módszerek fent vázolt gyakorlati alkalmazása felvet további két figyelemre méltó szempontot: a nyelvi analízist és a gépi reprezentációra fordítást végző algoritmus elképzelhető sebességének kérdése és az ebből következő gazdaságossági kérdések.

Tekintettel arra, hogy ha bármily hatékonynak is képzelünk egy nyelvi analízist és gépi reprezentációra fordítást végző algoritmust, aminek hatékonysága

egyébként nyilván fokozható előfeldolgozási, szűrési processzekkel, mégis maga a nyelvi analízis és a fordítás csak komplex logikai műveletként gondolható el, elemzési, illesztési és összehasonlítási műveletek halmazával, melyek időszükséglete az algoritmus fokozatos csiszolásával nyilván folyamatosan szorítható egy minimum felé, de ez a minimum a jelenlegi és a jövőbeni hardver-képességek mellett is mindenképpen jelentős érték marad. Eddigi tapasztalataink szerint, kielégítő teljesítmény lenne, ha egy átlagos összetett mondat kiértékelését az algoritmus 0,001 sec körül el tudná végezni. Ez azt jelenti, hogy egy átlagos könyv (300 oldal) „átolvasás" az algoritmusnak 2,5-3 másodpercet vesz igénybe. Egy ilyen eredmény már mindenképpen használhatónak mondható, mert pl. mentesítheti az embert attól, hogy fölöslegesen elolvassa a számára irreleváns irodalmat. Ugyanakkor ilyen módszerrel nekilátni egy könyvtárnyi anyag feldolgozásához egyetlen kérdés miatt, egyszerűen kilátástalan. (Az Országos Széchenyi Könyvtárban kb. 4,5 millió információs egységet tárolnak, aminek a túlnyomó többsége könyv.) Az ilyen nemzeti könyvtárban található információmennyiséget nagyságrendekkel meghaladó információtáraknak (mint az Internet) nekiesni hasonló módszerekkel pedig akkor is értelmetlen, ha a jövő ígéretébe, a kvantumszámítógépek világába képzeljük magunkat, több nagyságrenddel megnövelt számítási kapacitással.

Belátható, hogy nemcsak a korlátos gépkapacitás miatt, hanem az ismerethalmaz feldolgozásának gazdaságossága miatt is, csak különleges esetekben érdemes egy konkrét kérdés miatt egy teljes információs bázist analizálni. (Az Országos Széchenyi Könyvtár teljes anyagának elemzése egyetlen kérdés kapcsán a fenti adatokkal mintegy 140 napig tartana.) Az ilyen módszer rendkívüli gazdaságtalansága különösen akkor válik szembeötlővé, ha elképzelünk egy kérdést, amire végigfuttatjuk az analízist, majd felvetődik bennünk egy olyan kérdés, ami valamilyen szinten összefügg a már feltett kérdéssel, esetleg annak variánsa. Nyilvánvalóan sokkal gazdaságosabb lett volna a két egymással összefüggő kérdésre az analízist egyszerre futtatni, a szükséges gépidőt minimálisan növelve, szemben az egymás utáni futtatással, a gépidőt kétszeresére növelve. Ezen a nyomon elindulva gyorsan el lehet jutni odáig, hogy ha bizonyos kérdésekre van válaszunk, az bizonyos másik kérdésekre (pl. az egymást kizáró kérdésekre) a futtatást fölöslegessé teheti. Innen pedig már csak egy lépés az ismertbázis online faggatásának feladása, és valamilyen formalizált kérdések rendszerében rejlő lehetőségek szisztematikus kialakításának módszertani kutatása.

Némi megfontolás után azonban kiderül, hogy komoly kutatásra itt egyáltalán nincs szükség, mert a szóban forgó formalizált kérdések halmaza lényegében azonosnak tekinthető a szóban forgó háttértudás egy részével: az ontológiák, tezauruszok, osztályozási rendszerek által tárolt, rendszerezett fogalmakkal. Vagyis, a leghatékonyabb, ha veszünk egy jól rendszerezett fogalomtárat, és az algoritmusunkat nem alapvetően egy-egy ad hoc kérdés szolgálatába állítjuk, hanem az ismeretbázisunkat e fogalomtár szerint is rendezzük. Ez a módszer elveszti a tetszőleges kérdés feltételének közvetlenségét, de a keresést lényegében a felhasznált fogalomtár logikai hálóján hatékonyan és rendkívüli gyorsasággal végezhető

keresési műveletre vezeti vissza.

Természetesen, az ismerethalmaz és az n elemű fogalomtár szerinti rendezése elvben azt jelenti, hogy az n kérdést a teljes ismerethalmazon végigfuttatva előáll az az $m(< n)$ elemű szignifikáns fogalomtár, amely a szóban forgó ismerethalmazt pontosan jellemzi. Ez annak ismeretében válik fontossá, hogy pl. az ETO középkiadása mintegy 80.000 nyelvi egységet tartalmaz, ami az implicit információk figyelmen kívül hagyása esetén, a teljes ismeretbázison 80.000 kérdés végigfuttatását feltételezné, a fenti alapadatokat figyelembe véve mintegy 30.000 év időszükséglettel.

7. Összegzés, jövő, feladatok

A cikkben vázoltuk a könyvtári információkeresés elvi és gyakorlati hátterét különös tekintettel a számítástechnikai eszközök felhasználására. A létező kereső módszerek két egymástól lényegesen különböző szempontot megvalósító családba sorolhatók, úgymint a hagyományos könyvtártudomány vonalát követőek, illetve a gépi számítástechnikai szemléletűek. Előbbi reprezentálja a hosszú idő alatt felgyűlt specifikus könyvtártudományi ismereteket, és tartalmi keresésnek is nevezhetjük, utóbbi alkalmazza a számítógépek nyers technikai és algoritmikus képességeit, és mint ilyen a formai keresésnek is tekinthetjük. Sajnálatos tény, hogy e két megközelítés sokáig távol állt egymástól, konfliktusossá téve a kapcsolatot a két terület szakértelme között. A probléma a 90-es évektől kezdve tudatosult e határterület művelői között, és számos megoldási kísérlet született a tartalmi szempontokat adekvát módon kiszolgáló számítástechnikai eszközök megalkotására. E módszerek a mai napig nem érték el azt a szintet, amely egy gördülékeny kereső-eljárás megalkotásához nélkülözhetetlennek látszik. A jelen megoldások fő hiányossága a szükséges jelentős manuális előkészítő munka, köszönhetően a természetes nyelvű szövegek és a gépi ábrázolásuk közti lényeges különbségnek.

Jelen cikkben vázoltunk egy alternatív lehetőséget, mely a természetes nyelvű szövegek automatikus tartalmi/logikai leképezését képes megvalósítani, illetve ilyeneken keresést, következtetést végezni. Lehetőség nyílik a létező tudásanyagok integrálására, az azokon történő navigálással együtt. Az élő gyakorlatra tekintettel elemeztük a tartalmi keresőrendszerek hatékonyság problémáját is, ami a különböző rendszerek egymást kiegészítő, párhuzamos alkalmazásának fontosságára mutat rá.

A hivatkozott szövegreprezentációs módszer jelenleg fejlesztési fázisban van. Implementációja a feladatok széles köréhez adhat alkalmas eszközt. Közvetlen célként a létező könyvtári keresőrendszerekbe való automatikus szövegbesorolás képességét céloztuk meg. Később specifikus ismeret-anyagra vonatkozó szakértői rendszer kiépítését tervezzük. Távlati célként többnyelvű tudásháttérre alapozott, rugalmas ember-gép kérdezz-felelek kommunikációt céloztunk meg. Alkalmazási területként olyan tudományágak kerülhetnek szóba, amelyek világos, egy-

értelmű, rögzített fogalomrendszerrel fejtik ki tárgyukat, pl.: jog, orvostudomány, mérnöki tudományok, stb.

Hivatkozások

[1] Bálya D.: Az informatika kihívása a teszt-technológiában, BME TIO, 1997.

[2] Dán K., Haralyi E.: Könyvtárhasználati ismeretek a kerettantervben, http://www.om.hu/letolt/kozokt/konyvtar.doc

[3] Fülöp G.: Az információ, Kriterion, Bukarest 1990.

[4] Iwańska, L. M., Shapiro, S. C. eds.: Natural Language Processing and Knowledge Representation. Cambridge, MIT Press, 2000.

[5] Kunszt Gy. A tudományos kutatás logikai modellezése és tematikai irányítása, Akadémiai Kiadó, Budapest, 1975.

[6] Rédey G.: iCTRL: Intensional conformal text representation language, *Artificial Intelligence*, **109**/**1-2**(1999)33-70.

[7] Ruzsa I.: Logikai szintaxis és szemantika, 2. köt. Akadémiai Kiadó, Budapest, 1988.

[8] Sowa, J. F.: Knowledge Representation: Logical, Philosophical and Computational Foundations, PWS Publ. Co., Pacific Grove, CA 1999.

[9] Ungváry R., Vajda E.: Könyvtári Információkeresés, Typotex 2002.

[10] Vincze T.: Hálózati kislexikon, http://gisfigyelo.geocentrum.hu/informatika/-kisokos_informacio.html