

HumInA projekcsoport a ReALIS1.1 bázisán

Nóthig László, Alberti Gábor

PTE BTK Nyelvtudományi Tanszék

ReALIS Eméleti, Számítógépes és Kognitív Nyelvészeti Kutatócsoport
nothig.laszlo@gmail.com, alberti.gabor@pte.hu

Kivonat: A laptopos bemutóra benyújtott szoftver egy mobiltelefonra szánt alkalmazás személyi számítógépes verziója. Létrehozásának fő célja a mindennapi kommunikációt meghatározó kapcsolati információk gyors áttekintése, menedzselése. A szoftver a ReALIS főbb felismerésein alapuló, annak szellemiségét hordozó, könnyen kezelhető alkalmazás, amelyhez hamarosan kapcsolódhat egy, a korábbi elméleti munkákon és számítógépes programjainkon alapuló nyelvi elemző, ahol a rendszer megkeresi az ismerhető szereplőket, entitásokat, eseményeket. A ReALIS rendszer HumInA – Humán Intelligenciájú Adatbázis – modulja egy potenciálisan széles felhasználói kör részére elkerülhetővé teszi azt a munkát, amit egy világmodell részletes kidolgozása jelent. Az elnevezés onnan ered, hogy az információt nem önmagában, hanem az emberi kommunikációs folyamattal együtt próbáljuk megragadni, hogy ezáltal további értékes adatokhoz jussunk. A bemutatásra kerülő szoftver legegyszerűbb funkciója különböző mondatok tárolása, egy használható és áttekinthető világocska-rendszer létrehozása, valamint a más-más kontextusban előforduló entitások és események közötti kapcsolatok rögzítése. A mindennapi életben általában másoktól szerzett információra támaszkodunk, ahol forrásaink is csak részleges tudással rendelkeznek. Amikor egy kijelentés igazságáról szeretnénk dönteni, valamilyen súlyozás szerint vesszük figyelembe a környezetünkől érkező információkat. A program különböző stratégiák alapján megbízhatósági kereséseket tud végrehajtani, ahol eltérő heurisztikák alkalmazásával próbálja figyelembe venni az egyes források hitelességét.

1. Az információtartalom új dimenziói

1.1. Már nem az arany, a vas, a szén vagy a kőolaj megszerzése, birtoklása, továbbítása és elosztása mozgatja a világot, hanem az információ vált a jelenkor legfontosabb „ásványkincsévé”. Olyan mennyiségű adat vált elérhetővé az interneten, ami csodálatos lehetőségeket kínál egyfelől, másfelől viszont ott a veszély, hogy belefutunk, elveszünk benne.

A területtel nyelvészeti és informatikai kongresszusok külön szekciói foglalkoznak, mi most egy új megközelítésre hívjuk fel a figyelmet.

Nyilvánvaló, hogy adatbázisainkat a lehető legintelligensebb módon kell megszervezni... Mi azonban ez a „legintelligensebb mód”? Az emberi intelligencia?

Aligha választhatunk más kiindulópontot. A legóvatosabb megfogalmazás mellett is kijelenthetjük, hogy érdemes a humán intelligenciából a lehető legtöbbet ellesni, és

az eltanult elemeket kamatoztatni az adatbázis-szervezés területén. Mintha az eddigi adatbázisokat „kommunikációs tudatra” ébresztenénk.

1.2. A nyelvtudományi háttérben a döntő felismerés az, hogy egy-egy mondat információtartalma nem pusztán a világban értékelhető logikai tény (*igaz vagy hamis?*). Hanem az emberi kommunikációban megsokszorozódik, bizonyos értelemben újabb dimenziókat nyer [1], és az ezekben a dimenziókban rejlő információ gyakran fontosabb, mint a „csupasz tények” [7]. Ha nem ragadjuk ki az információt a kommunikációs folyamatból, hanem azzal együtt ragadjuk meg, akkor a kommunikációs folyamat úgy működik, mint egy prizma, amelynek sok-sok lapján megsokszorozva és némileg mutálódva észleljük az elemi információs egységeket.

1.3. Vegyük górcső alá az (1a) példában megadott egyszerű mondatot:

1. példa. Információ, túl az igazságon...

- a. Anna tegnap felhívta Bélát.
- b. Anna tegnap felhívta Bélát?
- c. Dóra a kérdés elhangzását követően tisztában van azzal, hogy
Eszter megtudta:
Csaba úgy gondolja, hogy
ő, Dóra, tudja, hogy
felhívta-e tegnap Anna Bélát.
- d. Anna egy adott napon felhívta Bélát.
- e. Mire vágyik Csaba, hogy
Eszternek milyen kérdésekben legyen tévhitelme?

Ha kérdő mondatként hangzott el (1b), akkor még csak az sem derül ki, hogy igaz-e az említett hívás ténye, vagy sem. Vegyünk tehát ilyen jellegű esetet: tegyük fel, hogy Csaba éppen ezt kérdezte Dórától, Eszter jelenlétében! Mi ebben a nyelvileg megragadható információ?

1. Az, hogy Csaba a kérdéssel kinyilvánítja, hogy ő nem tudja, hogy Anna tegnap felhívta-e Bélát. 2. Dóráról viszont azt feltételezi, hogy ő tudja az igazat. 3. Túl a tudáson és feltételezésen, vágyak és szándékok is megfogalmazódnak. Csaba például a szóban forgó kérdéssel kinyilvánítja azt a vágyát, hogy szeretné tudni, felhívta-e Anna Bélát. 4. Csaba szándéka: felébreszteni Dórában a szándékot kíváncsisága kielégítésére. 5. Mindezt (1–4.) érdemes lehet regisztrálni egy adatbázisban, például egy nyomozás vagy egy bírósági tárgyalás bármelyik szereplője számára, megragadva így módon Csaba (adott pillanatbeli) „információállapotát”. Arról sem megfeledkezve az adatbázis-építés során, hogy Dóra és Eszter is levonhatta az 1–4. következtetéseket, hosszabb távon akár olyan összetett gondolatokra jutva, amelyet a fenti (1c) pontban említünk.

A fenti példával azt kívántuk bemutatni, hogy érdemes olyan adatbázist építeni, amelyben a fenti (1d) pontban megadott információ mellett nem pusztán egy IGAZ vagy HAMIS értékelés áll, hanem egy rugalmasan bővíthető hipotézisegyüttes egymással kommunikációs kapcsolatban álló szereplők tudásáról, vágyairól, szándékairól. Az is értékes információ, hogy valaki nem IGAZ vagy HAMIS választ adna egy kérdésre, hanem valamiféle „0” igazságértéket („nem tudom”).

Szögezzük le gyorsan: a regisztrált adattömeget nem kell készpénznek venni. Csaba akár „hazudhatott” is a kérdésével, amennyiben például valójában biztos tudása van arról, hogy Anna felhívta Bélát. Ez a fajta mímelés érvénytelenné teszi a fenti 1–4. pontokban regisztrált információ egy részét. Ez azonban egyáltalán nem jelenti azt, hogy az információ rögzítése fölösleges vagy hibás lépés volt. Csupán azt jelenti, hogy az adatbázis felhasználójának fel kell készülnie arra, hogy egyes adatcsoportokra átértékelő műveleteket kell majd alkalmaznia, bizonyos eséllyel. Amennyiben például biztosan tudja, hogy Csaba tisztában van a szóban forgó telefonhívás létevel, rendkívül fontos információhoz jut: ahhoz, hogy Csaba meg akarja tévesztetni a megkérdezett Dórát és/vagy a beszélgetésüket figyelő Esztert. Egy intelligensen strukturált adatbázison alternatív értékelési mechanizmusokat lehet lefuttatni, a felhasználói igények függvényében. Vagy szűrni lehet bizonyos adat-típusokra, például a fenti (1e) pontban megadottra (pl.).

2. Felhasználói adatbázisok építése

2.1. A címben említett HumInA projekt keretében ilyen jellegű *humán intelligencia szerint szervezett adatbázisokat* építünk ki a ReALIS1.1 keretrendszerre alapozva [2] [6].

Az elmúlt években a ReALIS [1] számítógépes implementációja során két eltérő, de egymást kiegészítő megközelítés merült fel a nyelvi elemzés megvalósításakor. Az egyik esetben a vonzatokra és az egyéb szintaktikai összetevőkre fókuszáltunk, hogy az adott (magyar) nyelv szabályait vezérfonálnak választva építsük fel a mondatot. A másik – „nyelvfüggetlen” – esetben elképzelésünk szerint a mondat elemeinek összerendelése eleinte a felhasználó feladata kell, hogy legyen – bár a program természetesen ezt több eszközzel is támogatja. Ha az elsődleges cél a mondatok igazságértékelése, hosszabb szövegek pragmatikai-szemantikai elemzése – ami egyébként a valódi kihívást is jelenti –, akkor a második megközelítés is vállalható, hiszen a szintaktikai elemzés „csak” ahhoz szükséges, hogy a mondatban megtaláljuk a – megfelelő – kapcsolatokat az egyes összetevők között.

A ReALIS rendszer HumInA modulja a saját területén feloldja ezt a dilemmát, és egy potenciálisan széles felhasználói kör részére elkerülhetővé teszi azt az egyébként nem megspórolható munkát, amit egy világmodell aprólékos kidolgozása jelent.

Az alkalmazás legegyszerűbb funkciója különböző mondatok tárolása, amelyeket egymástól eltérő nézetekben láthatunk. Az alapképernyő egy munkaasztal, amely megjeleníti az adatbázis mondatait, a respektált szereplőket, ismerősöket, a mobil tulajdonosát és a szereplők viszonyát az állításokhoz, vagyis melyiküknek milyen világocskája, vagy világocskái „tartalmazza” (tartalmazzák) az adott információt. A megvalósítás egyik nehézsége egy többdimenziós halmaz két dimenzióba való ergonomikus leképezése volt. Ez egy idő után ahhoz az immár technikai kérdéshez vezetett, hogy miként lehet az egyik „összevont” dimenzió mentén az információt könnyen áttekinthetővé és gyorsan bejárhatóvá tenni. Bár már egy tucatnál alig több „ismerős” és 3 szintű világocskacímke-rendszer esetén is több ezer az egy mondatához tartozó világocskacímke-helyek száma és ezek mindegyike több világocskacímke

„tartalmazhat”, a kész alkalmazásban – ennél sokkal több szereplő esetén is – a navigálás könnyen és gyorsan történik.



1. ábra. A program áttekintő képernyője.

2.2. A program használata során a felhasználó egy karaktersorozatot – általában egy mondatot – ír az adatbázisba. Ezt olyan mélységig elemzi, amilyen mélységig akarja. A bejegyzés egysége az infon [7], azaz valamilyen eseménybe való „belelátás” egy kiválasztott időpillanat aspektusából [1,3]. Magát az infont a rendszer nem elemzi, viszont rögzíthetők hozzá a felhasználó számára releváns szereplők, a további entitások és a kapcsolódó események. Az egyes elemek összekapcsolhatók, hierarchiába rendezhetők. A felhasználó maga dönti el, hogy egy esemény ugyanaz-e, mint ami egy korábbi bejegyzésben szerepelt, esetleg az esemény egy más aspektusában (pl. korábban mint egy jövőbeli terv jelent meg, most pedig éppen a kumulatív szakaszában tart [3]). Létre lehet hozni forgatókönyveket (definiálható címkével ellátott űrlapokat, rendezett n-esekből álló sablonokat) az egyes eseményekre, amelyekkel kapcsolat létesíthető az események és a szereplők között. Az infonok különböző világocskákba – a külvilágnak az egyes szereplők elméjében a tudás, a vágyak, a szándékok szerint megsokszorozott alternatív világába – kerül(het)nek, amelyek a „Tudom, hogy Jóska tudja, hogy én tudom...” állításhoz hasonló kijelentésekhez vezetnek. A világocskákban lévő információ áttekinthető struktúrákban jelenik meg, a bevitt adatok és kapcsolataik különféle nézetekben láthatók, a rendszerben különböző lekérdezések és listák készíthetők.

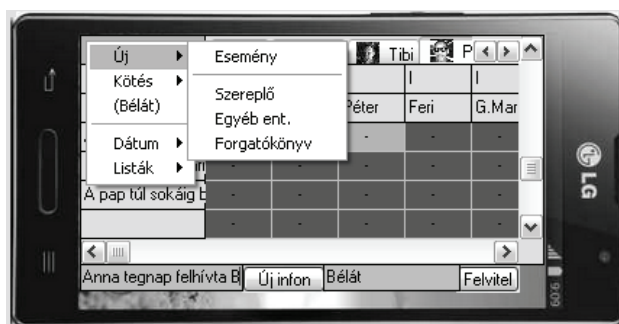
2.3. A fentieket is szeretnénk megvilágítani egy példával. Tekintsük a következő két mondatot [4]:

2. példa.

- a. Mari múlt szombaton feleségül ment Péterhez.
- b. ... A pap másfél órán át beszélt a házasság nehézségeiről.
- b'. ... ??A kutya veszettül ugatott.

Ebben a klasszikus példában az első mondatból bár logikailag nem következik a pap jelenléte, kulturális enciklopédikus tudásunkban az esküvőhöz asszociálódik $(2a+b)$, szemben egy kutya említésével $(2a+b')$.

A felhasználó a mondatokat minden további információ megadása nélkül is beírhatja, így azonban csak jegyzet írására használja a programot. Ha azonban bármilyen okból úgy gondolja, hogy érdemes más adatokat is rögzíteni (pl. több Péter nevű ismerőse van, vagy valamelyik szereplő később várhatóan újra elő fog fordulni az adatbázisban), megteheti a következőket: Felvesz egy Péter nevű entitást, mint új személyt, vagy azonosítja egy korábban már felvett szereplővel (pl. *az egyetlen unokatestvérem*). Mari szintén bekerülhet, mint új szereplő, vagy azonosítható egy korábbival. Nemcsak az entitásokat, hanem az eseményt is "nevesíthetjük", létrehozhatunk egy új eseményt (vagy ha erre már korábban sor került, ahhoz hozzáköthetjük), aminek címkéje pl. *Mariék esküvője*. A második mondatot már ehhez az eseményhez tudjuk kötni. Ezek után a *pap* szintén felvehető, mint (valószínűleg új) személy. Az esküvő általában sok szereplő részvételével zajlik, amelyre előre megírt forgatókönyv készülhet, és később az aktuális esemény sablonjaként használható. Ezen az űrlapon lesz tehát egy menyasszony, egy vőlegény és további rovatok esetleg opcionális szereplők – például a *pap* – számára. Ezt a szereposztást a sablon alapján könnyen elkészíthetjük a beírt mondatokhoz az adott eseményre. A tipikus forgatókönyvek megírása nem szükségképpen a felhasználó feladata, célszerűen azok az alkalmazás letöltésekor, frissítésekor kerülhetnek a mobil eszközre.



2. ábra. Új szereplő felvitele.

2.4. A sokszor látszólag magától értetődő kapcsolatok létrehozásakor döntő a felhasználóbarát működés – különösen egy mobiltelefonos alkalmazás esetén. Ennek elérésére szolgál például a kiválasztott szó kattintással történő szerkesztőterületre hozása, az aspektus menü nélküli kiválasztása, vagy a megjelenítésben annak az elvnek a végigvitele, hogy a mondatrészek soha nem önmagukban, hanem előfordulásukkal, teljes környezetükkel együtt jelennek meg. A munkaasztal fő vizuális komponense egy csúszka segítségével a rácsméret változtatásával az olvashatóság határáig „fokozatmentesen” átméretezhető, de a kurzor mellett minden esetben olvasható az alatta lévő cella, vagy mező tartalma.

2.5. A megfelelő kötések létrehozása hasznos lehet a felhasználó számára, ugyanakkor modellezi az információ tárolását a mindennapi kommunikáció során. Bár a háttér lényegében ugyanaz, mint a ReALIS alaprendszerénél, akadnak eltérések is. Az alkalmazással a mobilkészüléket készségi szinten használókat céloztuk meg:

teljes világmodell építésre nincs szükség, nem „igazságértékelés” a cél. A beszélő és a hallgató (sőt további jelenlévők is) viszont ugyanúgy rögzíthetők és akár az alapmunkaasztalon is megjeleníthetők. Ez egy, a világocska-rendszer ábrázolásához képest elemibb információt tartalmazó nézet, amelynek szerkezete, keresetősége ugyanolyan. Ebből a nézetből a felvitt mondatok, pl. (1b) megfelelő világocskába történő „szétosztása” egy menüpont meghívásával történik, ahol a funkció végrehajtása előtt a felhasználónak be kell állítani a főbb kommunikációs szerepeket (a példánál elsősorban a kérdezőt). Egyáltalán nem mindegy ugyanis – amint láttuk – hogy a prototipikus alapkérdés hangzik el, vagy a korábban említett mímelésről, esetleg tanári vizsgakérdésről, netalán nyomozói keresztkérdésről van szó. A kialakuló adatbázis többféleképpen lekérdezhető. Egy magától értetődő használati mód az egyszerű keresés, ahol már nemcsak szótöredékes egyezés alapján fogunk találatokat kapni, hanem az egymással azonosított, ill. az asszociált entitások és események is az eredménylistára kerülnek.

2.6. A mindennapi életben az információ kis része jut közvetlenül, érzékszerveinken keresztül tudomásunkra. Általában másoktól szerzett adatokra támaszkodunk, ahol forrásaink is csak parciális tudással rendelkeznek. Amikor egy kijelentés igazságáról szeretnénk dönteni, figyelembe vesszük a környezetünkben érkező – lehetőleg mértékadó – információt. A program megbízhatósági kereséseket képes végrehajtani különböző stratégiák szerint. A figyelembe vett szereplők megfelelő világocskáik alapján kiértékeli az ugyanarra az állításra vonatkozó véleményeket, és elfogadásra vagy elutasításra javasolja a felhasználó számára. Ahhoz, hogy súlyozni lehessen az „informátorokat”, különböző eljárások készültek. A legegyszerűbb, amikor ismert igazságértékű állításokon teszteljük a szavazatokat. Más esetben különböző heurisztikák alkalmazhatók, amelyek jellemzően több iterációs lépést tartalmaznak.

Köszönetnyilvánítás

A jelen tudományos közleményt a szerzők a Pécsi Tudományegyetem alapításának 650. évfordulója emlékének szentelik. A szerzőket e cikk alapjait jelentő kutatásaikban és a konferencia-részvételben a TÁMOP 4.2.2.C-11/1/KONV-2012-0005 (Jól-lét az információs társadalomban) kutatási projektum támogatta.

Hivatkozások

1. Alberti, G.: *ReALIS: Interpretálók a világban, világok az interpretálóban*. Akadémiai Kiadó, Budapest (2011)
2. Alberti, G., Nöthig, L.: *ReALIS1.1: The Toolbox of Generalized Intensional Truth Evaluation*. In: Grzymala-Busse, J., Schwab, I., eds.: *INTELLI 2014, The Third International Conference on Intelligent Systems and Applications (INTELLI) (2014)* 60–66

3. Farkas, J., Ohnmacht, M.: Aspect and Eventuality Structure in a Representational Dynamic Semantics. In: Alberti, G., Kleiber, J., Farkas, J.: *Vonzásban és változásban*, PTE Nyelvtudományi Doktori Iskola, Pécs (2012) 353–379
4. Kálmán, L.: Deferred Information: The Semantics of Commitment. In: Kálmán, L., Pólos, L., eds.: *Papers from the Second Symposium on Logic and Language*, Akadémiai Kiadó, Budapest (1990) 125–157
5. Nőthig, L., Alberti, G., Dóla, M.: \Re ALIS1.1. In: Tanács, A., Varga, V., Vincze, V., eds.: *X. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, SZTE (2014) 364–372
6. Seligman, J., Moss, L.: Situation Theory. In: van Benthem, J., ter Meulen, A., eds.: *Handbook of Logic and Language*, Elsevier, Amsterdam / Cambridge (1997) 239–309
7. Vadász, N., Alberti, G., Kleiber, J.: The Matrix of Beliefs, Desires and Intentions. *International Journal of Computational Linguistics & Applications* 4/1 (2013) 95–110