

Bátfai Norbert – Bátfai Mária Erika

Virtuális könyvtáros segítheti majd a kutatókat kézírataik beküldésében a Debreceni Egyetemen

Jelen munkánkban egy magyar nyelvű, publikálást segítő AIML (Artificial Intelligence Markup Language) tudásbázis létrehozásának bemutatását kezdjük meg. Speciálisan egy informatikai kéziratok beküldését segíteni képes tájékoztató könyvtáros tudásbázisának létrehozására szűkítjük erőfeszítéseinket. Távlabbi célunk a kifejlesztett tudásbázis üzem során szerzett tapasztalatok analitikai elemzése.

Bevezetés

Joseph Weizenbaum „Eliza” programja [1] volt az a mag, amiből napjaink több száz csevegőrobotja származtatható. Eliza képes volt elhiteni a vele folytatott párbeszédéről, hogy azok nem közte és egy ember között, hanem két ember: egy pszichológus és a páciense között zajlottak le. A párbeszéd részletei összemérhetők [2] lehetnek a Turing-teszttel [3], amelynek során a teszten sikeresen átmenő programoknak azt kell elhítenniük a szakértő megfigyelőkkel, hogy ők nem programok, hanem valódi személyek. Folyamatosan fenntartja a csevegőrobotok iránti érdeklődést az 1991 óta évente átadott Loebner-díj [4], a Turing-teszt gyengített változata. A 2000., 2001. és a 2004. évben Richard Wallace „Alice AIML” (A.L.I.C.E. = Artificial Linguistic Internet Computer Entity; AIML = Artificial Intelligence Markup Language) alapú csevegőrobotja nyerte el a díjat [5].

Jelen munkánkban egy magyar nyelvű, publikálást segítő AIML tudásbázis létrehozását kezdjük meg. Speciálisan egy informatikai kéziratok beküldését segíteni képes tájékoztató könyvtáros tudásbázisának létrehozására szűkítjük erőfeszítéseinket. Célunk a tudásbázis elkészítése, majd üzemeltetése, s a későbbiekben az üzem során szerzett tapasztalatok analitikai elemzése, például a Zipf-törvény teljesülésének vizsgálata [6]. (A Zipf-törvény azt a megfigyelést rögzíti, hogy egy természetes nyelvi korpuszban a szavak gyakorisága a Pareto-eloszlást követi, azaz kevés szó nagyon sokszor szerepel és nagyon sok más szó csak ritkán fordul elő.)

Háttér, korábbi kapcsolódó munkák

Számos helyen találkozhatunk az AIML alkalmazásával. Kimondottan virtuális könyvtárosok létrehozásáról olvashatunk például a [7] és a [8] munkákban, online chatbot könyvtáros Kornélia [9] és BoB [10]. Találunk AIML-alapú csevegő entitásokat az iparban [11], az oktatás területén [12], vagy akár a SecondLife-ban [13] is. Sőt, magyar nyelvű csevegőként Lafoxkát [14] említhetjük. A *chatbots.org* portálon csevegőrobotok rövid bemutatásának nagy gyűjteményét találhatjuk meg, köztük több virtuális könyvtárossal. Ezek tipikus felhasználói felülete szöveges alapú, amit a csevegőrobot avatarját animáló képek színesítenek.

Érdekes trend, hogy a szóban forgó csevegőprogramok együttműködnek más természetes nyelvű feldolgozásalapú kutatási erőfeszítésekkel, például a *WordNet* vagy a *Cyc* adatbázisokkal [15, 16]. Az AIML értelmezőket betűkkel számozzák, nagy ugrás volt a Java-alapú Program D. Jelen munkánkat a Program W és a Python-alapú Program Y értelmezőket használva kezdjük el. Mindkettő nyílt forráskódú, előbbit GPL, utóbbit BSD licenccel látták el. Érdekességként jegyezhetjük meg, hogy a Program N össze van kapcsolva a ConceptNet tudásbázissal [17]. A ConceptNet háttérét adó Open Mind Common Sense [18] magyar lapjait magunk fordítjuk [19]. Az említett WordNet, Cyc vagy ConceptNet természetes nyelvi szemantikus tudásbázisok a számítógép és az ember érdemi természetes nyelvi interakciója irányába tett erőfeszítések.

Eredmények

A fejlesztés alatt álló, publikálást segítő AIML-alapú tudásbázis jelen pillanatban néhány általános célú állományból, mint a bemutatkozást, búcsúzást leíró *udvariassag.aiml.xml*, vagy a robot önmagáról alkotott képét leíró *robot.aiml.xml* fájlból áll. A *konyvtar.aiml.xml* állomány tartalmazza a könyvtári interakcióval általában kapcsolatos ismeretek leírását. Végezetül külön leíró fájlba kerül minden olyan informatikai szakfolyóirat, amelynek kapcsán tudással látjuk el a csevegőrobotot; ilyen például a TMT esetén a *tmt.aiml.xml* állomány. A szóban forgó állományok a www.inf.unideb.hu/~nbatfai/kk/ címről tölthetők le.

Az AIML-alapú csevegés lényegét a témákba szervezhető kategóriák jelentik, a tudás egysége a kategória. Az AIML-alapú értelmező legegyszerűbb válasza a beérkező kérdés és a tudásbázisban lévő kategóriák mintáinak összevetésén alapszik. A mintapéldánkból kiragadva tekintsünk erre egy esetet: tegyük fel, hogy a "HOGY PUBLIKÁLHATOK A JUCS FOLYÓIRATBAN" című téma már kijelölésre került és az alábbi kérdés érkezik.

Olvasó> Van publikációs díj?

Ez a kérdés az alábbi kategória mintájára (pattern tag) illeszkedik.

```
<category>
  <pattern>VAN PUBLIKÁCIÓS DÍJ</pattern>
  <template>
    <srail>A JUCS KÉR PUBLIKÁCIÓS DÍ-
      JAT</srail>
  </template>
</category>
```

Aminek válaszmintája (template tag) az *srai* rekurziót építő taggal azt mondja, hogy ugyanazt kell rá válaszolni, mintha csak „A JUCS kér publikációs díjat?” kérdésre válaszolnánk, amelyre pedig a következő kategória adja meg a választ.

```
<category>
  <pattern>A JUCS KÉR PUBLIKÁCIÓS DÍ-
    JAT</pattern>
  <template>
    Nincs publikációs díj.
  </template>
</category>
```

Itt a válaszmintában már nincs további rekurzió, a reakció tehát az lesz, hogy „Nincs publikációs díj.”

Olvasó> Van publikációs díj?

Robot> *Nincs publikációs díj.*

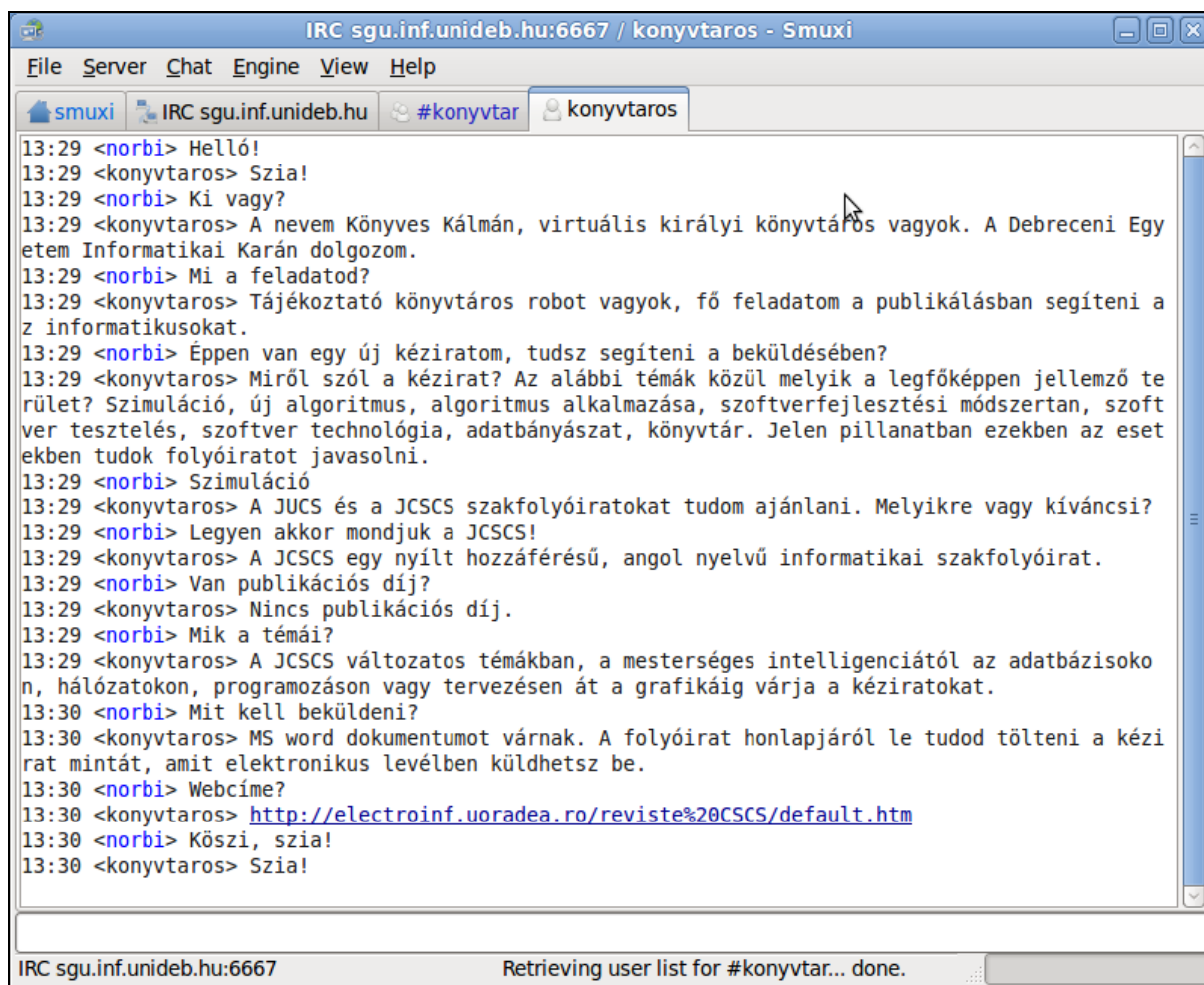
A *konyvtar.aiml.xml* állomány kategóriái próbálják úgy irányítani a beszélgetést, hogy az olvasó el tudja dönteni, hogy a felkínáltak közül melyik szakfolyóíratra kíváncsi éppen. Az ezzel kapcsolatos további kategóriák a kiválasztott kiadványra vonatkozó állományban találhatóak.

A tudásbázis tesztelését jelenleg a W és az Y értelmezőkkel végezzük. A W program esetén a robot könyvtárossal folytatott privát csevegést láthatjuk az *1. ábrán*, illetve ennek átirata olvasható az *1. táblázatban*. A D értelmező lehetőséget ad webalkalmazás készítésére, így a csevegőrobot könnyen integrálható az adott könyvtár weblapjába is. Tesztelési célokra gyors és kényelmes a PyAIML (Program Y) Python-implementáció használata. Az érdeklődő könyvtárosoknak, olvasóknak nem kell mást tenniük, mint a projekt *Sourceforge.net* lapjáról [20] letölteni a tömörített állományt, kicsomagolni, majd a kicsomagolt *aiml* könyvtárral egy szintre lépve indítani a *KiralyiKonyvtaros.py* szkriptet. GNU/Linux rendszer alatt ez a néhány egyszerű osztályból álló program az *espeak* beszéd szintetizátor használatával fokozza az érdeklődők felhasználói élményét. (Az említett Python teszt szkript része a fejlesztett AIML tudásbázisnak. A futtatásáról készített videó megtekinthető a YouTube videomegosztón a <http://www.youtube.com/watch?v=nVneMJt0UEo> címen.)

Értékelés

Az egyelőre szűk körben tesztelt készülő tudásbázissal biztatóak a tapasztalataink. Megmutattuk, hogy kis energiabefektetéssel viszonylag hatékony, a gyenge MI (mesterséges intelligencia) látszatát kelteni képes online segítő robot alakítható ki. E cikk benyújtásával egy időben keressük a lehetőségeket a helyi éles teszttüzem kialakítására. Ezt követően egyetemi kollégáinkat kérjük majd meg, hogy publikálás előtt kérjenek segítséget Könyves Kálmántól is. Az oktatási gyakorlatban viszont már alkalmazzuk a robotot, az érdeklődő hallgatói közösség egy folyamatos teszttüzem fenntartását is meg tudja oldani. Ez nélkülözhetetlen ahhoz, hogy megfelelő mennyiségű adat keletkezzen, amivel az analitikai elemzés egyáltalán megindulhat.

További munkaként adódik a tudásbázis folyamatos fejlesztése, amiben számítunk könyvtáros kollégák bekapcsolódására. Fejlesztői szempontból a folyóirat-specifikus állományok nagyon hasonlóak,



1. ábra Beszélgetés menete Könyves Kálmán virtuális királyi könyvtárossal a Program W értelmezőt IRC csatornán keresztül használva

ezért ezek generálására érdemes egy „back end” jellegű, adminisztrátori réteg biztosítása. Felhasználói szempontból pedig az egyszerű gépelési hibák automatikus javítása nagyban fokozhatja a csevegőrobot használhatóságát.

Érdemes elgondolkodni a csevegőrobot felhasználásán az általános könyvtári tájékoztatásban (ilyen munkát végez a már említett Kornélia és BoB is) mind helyben, mind a nyitva tartási időn túl: az online katalógusban és a könyvtár honlapján. Könyves Kálmán segédkezet nyújthatna az általános könyvtárhasználati tudnivalók átadásához, a szolgáltatások (meg)ismertetéséhez (beiratkozás, kölcsönzés, hosszabbítás, gyűjteményjellemzők, keresés az OPAC-ban, adatbázisokban stb.) – azaz egy interaktív, „emberközeli” felülettel kapnánk. Ezzel megtehetnénk néhány lépést a könyvtári felületek akadálymentesítése felé is,

egyben a MINERVA-ajánlás [21] 4. (elérhető minden felhasználó számára), 5. (felhasználóközpontú) és 6. (interaktív) pontjainak teljesítését is megkönnyítenénk.

Az olvasó így egy (általunk felkészített) könyvtárost találna 24 órában, az év minden napján a könyvtárban. (További izgalmas távlatokat nyithat meg Könyves Kálmán és a helyi ILS összekapcsolásának lehetősége.)

Ha a könyvtárosok oldaláról közelítjük meg a csevegőrobot alkalmazását: semmi okunk nincs a féltékenységre, nem vagyunk lecserélhetőek rá; a robot folyamatos karbantartást igényel, illetve megkönnyíti a napi rutininformációk átadását – nyugodtak lehetünk, hogy akkor is megfelelő tájékoztatást adunk, amikor nem is vagyunk ott.

1. táblázat

A Könyves Kálmán virtuális királyi könyvtárossal folytatott beszélgetés átirata

13:29 <norbi> Helló!
13:29 <konyvtaros> Szia!
13:29 <norbi> Ki vagy?
13:29 <konyvtaros> A nevem Könyves Kálmán, virtuális királyi könyvtáros vagyok. A Debreceni Egyetem Informatikai Karán dolgozom.
13:29 <norbi> Mi a feladatod?
13:29 <konyvtaros> Tájékoztató könyvtáros robot vagyok, fő feladatomban segíteni az informatikusokat.
13:29 <norbi> Éppen van egy új kéziratom, tudsz segíteni a beküldésében?
13:29 <konyvtaros> Miről szól a kézirat? Az alábbi témák közül melyik a legfőképpen jellemző terület? Szimuláció, új algoritmus, algoritmus alkalmazása, szoftverfejlesztési módszertan, szoftvertesztelés, szoftvertechnológia, adatbányászat, könyvtár. Jelen pillanatban ezekben az esetekben tudok folyóiratot javasolni.
13:29 <norbi> Szimuláció
13:29 <konyvtaros> A JUCS és a JCSCS szakfolyóiratokat tudom ajánlani. Melyikre vagy kíváncsi?
13:29 <norbi> Legyen akkor mondjuk a JCSCS!
13:29 <konyvtaros> A JCSCS egy nyílt hozzáférésű, angol nyelvű informatikai szakfolyóirat.
13:29 <norbi> Van publikációs díj?
13:29 <konyvtaros> Nincs publikációs díj.
13:29 <norbi> Mik a témái?
13:29 <konyvtaros> A JCSCS változatos témákban, a mesterséges intelligenciától az adatbázisokon, hálózatokon, programozáson vagy tervezésen át a grafikáig várja a kéziratokat.
13:30 <norbi> Mit kell beküldeni?
13:30 <konyvtaros> MS word dokumentumot várnak. A folyóirat honlapjáról le tudod tölteni a kéziratmintát, amit elektronikus levélben küldhetsz be.
13:30 <norbi> Webcíme?
13:30 <konyvtaros> http://electroinf.uoradea.ro/reviste%20JCSCS/default.htm
13:30 <norbi> Kösz, szia!
13:30 <konyvtaros> Szia!

Köszönetnyilvánítás

A publikáció elkészítését részben a TÁMOP 4.2.1./B-09/1/KONV-2010-0007 számú projekt támogatta. A projekt részben az *Új Magyarország Fejlesztési Terv*en keresztül az *Európai Unió* támogatásával, az *Európai Regionális Fejlesztési Alap* és az *Európai Szociális Alap* társfinanszírozásával valósult meg.

Köszönjük *Drótos László* hasznos lektori tanácsait.

Irodalom

- [1] WEIZENBAUM, J.: ELIZA – a computer program for the study of natural language communication

between man and machine. Communications of the ACM, 9. köt. 1. sz. 1966. p. 36–45. DOI = <http://doi.acm.org/10.1145/365153.365168>

- [2] CSIRIK János: Gépi megértés. = Magyar Tudomány, 48. köt. 12. sz. 2003. p. 1486–1489.
- [3] RUSSELL, S. J. – NORVIG, P.: Artificial Intelligence: a Modern Approach. New Jersey, Prentice-Hall, Inc., 1995. p. 1020–1043. ISBN 0 13 103805 2
- [4] Home Page of The Loebner Prize in Artificial Intelligence. <http://www.loebner.net/Prize/loebner-prize.html> (letöltve: 2010. 11. 10.)
- [5] WALLACE, R. S.: ALICEBOT (2010.) <http://alicebot.blogspot.com/> (letöltve: 2010. 11. 10.)

- [6] WALLACE, R. S.: The Elements of AIML Style. (2003.)
<http://www.alicebot.org/style.pdf> (letöltve: 2010. 11. 10.)
- [7] CHRISTENSEN, Anne: A Trend from Germany: Library Chatbots in Digital Reference. 2007.
<http://www.tilburguniversity.nl/ticer/07carte/publicat/07christensen.pdf> (letöltve: 2010. 11. 10.)
- [8] NEBLETT, Vanessa – SHIVERS, Cassandra – THINGVALL, Nils – TSUI, Bobby: The Digital Access Architect: This new library role anticipates users' needs and wields the latest technology to respond. = Library Journal, 11. köt. 1. sz. 2005.
<http://www.libraryjournal.com/article/CA6277393.html> (letöltve: 2010. 11. 10.)
- [9] Kornélia. Kornhausbibliotheken, Bern, 2009.
http://www.kornhausbibliotheken.ch/index.php?option=com_wrapper&view=wrapper&Itemid=64 (letöltve: 2010. 11. 10.)
- [10] BoB. Free University of Bozen.
<http://web.inf.unibz.it/bob/?popup=yes&language=en> (letöltve: 2010. 11. 10.)
- [11] SANTANGELO, A. – AUGELLO, A. – SORCE, S. – PILATO, G. – GENTILE, A. – GENCO, A. – GAGLIO, S.: A virtual shopper customer assistant in pervasive environments. 2007.
<http://www.springerlink.com/content/t21u64469u783753/fulltext.pdf> (letöltve: 2010. 11. 10.)
- [12] FONTE, Fernando A. Mikic – BURGUILLO-RIAL, Juan C. – NISTAL, Martín Llamas: TQ-Bot: An AIML-based Tutor and Evaluator Bot. = J. UCS, 15. köt. 7. sz. 2009. p. 1486–1495.
http://www.jucs.org/jucs_15_7/tq_bot_an_aiml (letöltve: 2010. 11. 10.)
- [13] REHM, M. – ROSINA, P.: SecondLifeR as an Evaluation Platform for Multiagent Systems Featuring Social Interactions: Demo Papers. 2008.
<http://portal.acm.org/citation.cfm?id=1402754> (letöltve: 2010. 11. 10.)
- [14] Lafoxk, <http://www.lafoxka.hu> (letöltve: 2010. 11. 10.)
- [15] COURSEY, Kino: Living in CyN: Mating AIML and Cyc together with Program N. 2004.
http://www.cyc.com/doc/white_papers/Cyn_description.pdf (letöltve: 2010. 11. 10.)
- [16] SZYMANSKI, J. – SARNATOWICZ, T. – DUCH, W.: Semantic Memory for Avatars in Cyberspace. 2005.
<http://www.fizyka.umk.pl/ftp/papers/kmk/05-Avatars.pdf> (letöltve: 2010. 11. 10.)
- [17] MOSTAFA AL MASUM, Shaikh – MITSURU Ishizuka: Integrating Natural Language Understanding and a Cognitive Approach to Textual Emotion Recognition for Generating Human-Like Responses. 2006.
<http://www.miv.t.u-tokyo.ac.jp/papers/mostafa-ICCIT06-textual-emotion.pdf> (letöltve: 2010. 11. 10.)
- [18] Open Mind Common Sense.
<http://openmind.media.mit.edu> (letöltve: 2010. 11. 10.)
- [19] BÁTFAI Norbert: A Magyar JózanÉsz. 2009.
<http://www.inf.unideb.hu/~nbatfai/MagyarConceptNet.pdf> (letöltve: 2010. 11. 10.)
- [20] STRATTON, Cort: PyAIML (a.k.a. Program Y): A Python AIML Interpreter.
<http://pyaiml.sourceforge.net> (letöltve: 2010. 11. 10.)
- [21] Minerva: A kulturális honlap minőségi követelményei.
<http://www.mek.oszk.hu/minerva/html/dok/kulthon/apminkov.htm> (letöltve: 2010. 11. 10.)

Beérkezett: 2010. XI. 11-én.

