

# Yako: egy intelligens üzenetváltó alkalmazás nyelvtechnológiai kihívásai

Farkas Richárd, Kojedzinszky Tamás, Zsibrita János, Wieszner Vilmos

Szegedi Tudományegyetem, Informatikai Tanszékcsoport  
Szeged, Árpád tér 2.  
{rfarkas, zsibrita, wieszner}@inf.u-szeged.hu

## 1 Intelligens üzenetváltó alkalmazások

Az infokommunikációs technológiák fejlődésével egyidejűleg megfigyelhető az a tendencia, hogy az ember-ember kommunikációra egyre nagyobb arányban használunk szöveges formát (gondoljunk csak a Facebookra, chatre, e-mailre, SMS-re stb.). Számos szöveges kommunikációt támogató, üzenetváltó alkalmazás érhető el különféle platformokon, azonban ezek többsége nem rendelkezik „intelligens” funkciókkal [1]. Intelligens funkció lehet az üzenetek kategorizálása/csoportosítása, rangsorolása, összefoglalása, megjelölése vagy a szövegbevitelt támogató funkciók. Ezek a funkciók különösen hasznosak az okostelefonokon futó üzenetváltó alkalmazások számára, mert ott a kis megjelenítő- és beviteli eszközök miatt a hagyományos alkalmazások kényelmetlenek.

A Szegedi Tudományegyetemen, a FuturICT.hu projekt keretében kifejlesztettünk egy Android-alkalmazást (Yako), amely egységes felületen fogad és küld SMS-eket, e-maileket és alkalmas Facebook-beszélgetésre. Az alkalmazás felhasználóinak egy csoportjának folyamatosan mentjük, hogy hogyan váltanak üzeneteket ismerőseikkel (természetesen az adatvédelmi irányelvek tiszteletben tartása mellett<sup>1</sup>). Az így gyűjtött adatbázis jó alapul szolgál az üzenetváltásokhoz kapcsolódó intelligens funkciók fejlesztéséhez. A jelenleg fejlesztés alatt álló intelligens funkciók nagy része a nyelvtechnológia területéhez kapcsolódik: üzenetek fontosság szerinti rangsorolása, automatikus válaszgenerálás, témaszálak azonosítása, összefoglalás/kulcsszavazás, információk kiemelése. Az összegyűjtött adatbázis alapján megállapíthatjuk, hogy számos speciális nyelvtechnológiai kihívásokkal kell szembenéznie azoknak, akik magyar nyelvű szöveges üzenetváltást támogató intelligens funkciókat terveznek megvalósítani.

## 2 Ékezetesítés

Az első ilyen kihívás az automatikus ékezetesítés. Míg a desktopokon írt magyar üzenetek 95%-a ékezethelyesnek mondható, a mobil eszközökről írt üzeneteknél ez az arány mindössze 63%. A nyelvi előfeldolgozás egyik fontos lépése az ékezetek hely-

---

<sup>1</sup> A projekt keretében egy adatvédelmi hatástanulmány is elkészült.

reállítása [3]. A [2]-ben alkalmazott módszerekhez hasonlóan kidolgoztunk egy egyszerű eljárást, ami nagy ékezet helyes szótár segítségével, illetve többértelműség esetén a környező szavak figyelembe vételével valószínűségi döntést hoz.

### 3 Kérdések elemzése

Míg az üzenetváltások a szóbeli dialógusokhoz hasonlítanak, addig az elérhető annotált korpuszok leíró dokumentációkat tartalmaznak, melyekben jellemzően kevés kérdés fordul elő. Az üzenetváltások esetében nagyon gyakori a kérdés-válasz jellegű párbeszéd. A sztenderd korpuszokon tanított elemzők azonban a kérdéseken kifejezetten rosszul teljesítenek (mert kevés tanítópélda állt rendelkezésükre). Első funkcióként implementáltunk egy eldöntendő kérdéseket, illetve a döntés tárgyát azonosító egyszerű módszert. A kérdések elemzését egy gyorsválasz funkcióban használjuk fel. Egyszerű eldöntendő kérdés esetén *igen/nem* a lehetséges gyorsválasz, míg alternatívák közti választás esetén maguk az alternatívák. Például a *A TIK parkban vagy a Pivo Várban találkozunk?* kérdésre a *TIK parkban* és a *Pivo Várban* lesz a két gyorsválasz. A felismerő szabályok a magyarlanc [4] szófaji elemzésén alapulnak, így a *Hogy vagy barátom?* kérdésre nem lesz gyorsválaszi lehetőség a *Hogy* és a *barátom*.

### 4 Szövegen kívüli kontextus

Az intelligens funkciók megvalósításához nem elégséges a szöveges üzenetek elszigetelt elemzése, a tágabb értelemben vett környezet (pl. ki a partner) elemzése engedhetetlenül fontos. Az alkalmazásba bevezettünk ún. zónákat, amelyek közt a mobil eszköz a geolokáció alapján vált. Célunk az, hogy a rendszer különböző zónákban, különböző élethelyzetekhez igazodva másképp működjön. Különböző üzenetek fontosak például akkor ha otthon vagyunk, mint ha a munkahelyünkön tartózkodunk. A yako alkalmazásba építettünk egy szabályalapú fontosüzenet-detektáló módszert, ami a különböző zónákban különböző üzeneteket tekint fontosnak (a fontos üzenetek a beérkezett lista elejére kerülnek és az értesítésük is figyelemfelkeltőbb).

A bejövő üzenetek fontosságának eldöntésekor a szöveges tartalom mellett természetesen nagyon fontos a feladó személye és a fogadó és a feladó viszonya, mint szövegen kívüli kontextus. A rendszerünk egyszerű statisztikák felhasználásával egy felhasználó kapcsolatait öt különböző kategóriába sorolja be és ezeket a kategóriacímkeket felhasználjuk a fontosság eldöntéséhez.

### Köszönetnyilvánítás

A kutatás a futurICT.hu nevű, TÁMOP-4.2.2.C-11/1/KONV-2012-0013 azonosítószámú projekt keretében az Európai Unió és az Európai Szociális Alap társfinanszírozása mellett valósult meg.

## Hivatkozások

1. Dredze, M.: Intelligent Email: Aiding Users with AI. PhD thesis, Computer and Information Science, University of Pennsylvania (2009)
2. Zainkó, Cs., Németh, G.: Ékezetek gépi helyreállítása. In: A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó (2010) 485–488
3. Kornai, A., Tóth, G.: Gépi ékezés. Magyar Tudomány No.4 (1997) 400–410
4. Zsibrita, J., Vincze, V., Farkas, R.: magyarlanc 2.0: szintaktikai elemzés és felgyorsított szófaji egyértelműsítés. In: Tanács, A., Vincze, V. (szerk.): IX. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Szegedi Tudományegyetem (2013) 368–374