

Megértést segítő részletező gépi névfelolvasás magyar nyelvre*

Fék Márk¹, Németh Géza¹, Olasz Gábor^{1,2}, and Gordos Géza¹

¹ BME Távközlési és Médiainformatikai Tanszék,
1117 Budapest, Magyar tudósok körútja 2.

{fek,nemeth,gordos}@tmit.bme.hu

² MTA Nyelvtudományi Intézete,
Kempelen Farkas Beszédkutató Laboratórium,
1399 Budapest, Benczúr u. 33, Pf. 701/518
olaszy@nytud.hu

Kivonat Az automatikus beszédválaszú számszerinti tudakozó névfelolvasó modulja olvassa be a telefonba a keresett előfizető nevét. A felovastott személy- vagy cégnév telefonon keresztüli érthetőségének növelésére, a szótagoláshoz hasonló, részletező felolvasási móddal egészítettük ki a rendszert. A szótagokra bontás felteszi, hogy magyar nyelvű szöveggel van dolgunk. Idegen írásmód esetén, illetve ha a telefonon keresztüli gépi hang érthetősége nem megfelelő, az adott szótag után a megértést segítő megjegyzéseket iktat be a rendszer. A cikk ismerteti a rendszer felépítését és a megvalósítás során felmerült problémákat.

Kulcsszavak: gépi beszéd, névfelolvasás, számszerinti tudakozó, automatikus szótagolás

1. Bevezetés

Az automatikus beszédválaszú számszerinti tudakozók a gépi beszédkezelés egyik fontos alkalmazási területét képviselik [Spiegel, 1993], [Nebbia et al., 1998], [Lundin, 1998]. Az első ilyen magyar nyelvű rendszert a Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszéke és a T-Mobile Magyarország Rt. fejlesztette ki közösen [Németh et al., 2003]. A rendszer a Profivox magyar nyelvű beszéd szintetizátor [Olaszy et al., 2000] célirányos továbbfejlesztésével jött létre.

Ezekben a rendszerekben a felhasználó telefonon keresztül lép kapcsolatba az automatikus központtal, majd a nyomógombok segítségével megadja a keresett előfizető telefonszámát. A központ – ha adatbázisában megtalálta a telefonszámot, és az nem titkos – gépi beszéd segítségével beolvassa az ahhoz tartozó előfizető nevét és címét.

* A munkát a Nemzeti Kutatási és Fejlesztési Program, Alkalmazott beszédinformációs rendszerek projekt keretében támogatta

Gyakran előfordul, hogy az előfizető (magánszemély vagy cég) neve idegen, vagy rendhagyó helyesírású szavakat tartalmaz, így az ügyfélnek gondot jelenthet azok feljegyzése. Erre megoldást jelenthet az előfizető nevének betűzése (például: Techno='t, mint Tamás, e, mint Elemér,...'). A betűzés azonban lassú megoldás. Ennél gyorsabb, ha az idegen vagy rendhagyó helyesírású nevet szótagolva, és az egyes helyesírási sajátosságokat részletezve, olvassuk fel. Példaképp a 'techno-city' nevet 'teh, céhával, no, kötőjel, ci, ti, ipszilonnal'-ként olvassa fel a gép.

A fenti szótagoló-részletező megoldás hasonló a magyar anyanyelvű beszélők által gyakran használt eljáráshoz, mely segítségével szótagokra bontva magyarázzák el egymásnak egy idegen eredetű szó helyesírását. Ez a betűzésnél természetesebb megoldás várhatóan kedveltebb lesz a felhasználók bizonyos körében.

2. A részletező felolvasást végző modul felépítése

A részletező felolvasási feladatot a név- és címfelolvasó szoftver ún. részletező modulja valósítja meg. Ennek blokkvázlata az 1. ábrán látható. A bemenet a feldolgozatlan szöveg, a kimenet a részletezett, hanglejtési és beszédsszünet címkékkel ellátott szöveg. A feldolgozás lineáris, az egyes almodulok egymást követik. Az almodulok vagy lecserélnék egy adott szövegrészt, vagy kiegészítő megjegyzéseket szúrnak be. A beszűrt megjegyzések védelmet kapnak, így a további almodulok ezeket a szövegrészeket nem módosítják. A részfeladatokat szabályok és kivétel szótárak együttes alkalmazásával oldottuk meg. A kivételsszótárakat memória- és sebesség-hatékony ternáris keresőfákkal valósítottuk meg [Bentley and Sedgewick, 1997].

3. Magánhangzók és mássalhangzók címkézése

A részletező felolvasáshoz a szövegben bejelöljük a szótaghatárokat. E művelet első lépése a szöveg karaktereinek felcímkézése magánhangzó, mássalhangzó, szám, illetve egyéb karakter kategóriákra. A magyar nyelvben fonológiai kategória a beszédhang hossza. A mássalhangzók esetében a hosszú hangot az írásban betűkettőzéssel jelöljük. Ebből kifolyólag a betűk környezetét is vizsgálni kell, hogy a hosszú hangot egyértelműen azonosíthassuk a részletező felolvasás számára.

A magánhangzók és a mássalhangzók betűképeit egy-egy adatbázis tartalmazza. A magánhangzók között szerepelnek cég és személynevekben előforduló idegen betűk (például az ä), a mássalhangzók között a két betűvel jelzettek (sz, zs, stb.) és ezek hosszú megfelelői (ssz, zzs, stb.) is. Emellett a nevekben gyakran előforduló idegen mássalhangzók, mint például az sch, szintén szerepelnek.

A régies írású nevekben előforduló cz betűkapcsolat kezelésére az esetek jó részét lefedő szabályt alkottunk. Ha mássalhangzó után jön a cz, akkor egy betűnek tekintjük, egyébként különálló c-nek és z-nek. Így például a 'Kótczy' név 'kót, ci, cézéval, ipszilonnal'-ként részleteződik. Az olyan cz-t tartalmazó neveket, amelyek ez a szabály nem kezel, a későbbiekben működésbe lépő elválasztási

kivétel szótárban adtuk meg, így ezen szavak esetén a c és z között nem lesz elválasztás.

Külön problémát jelentett az y-os szavak kezelése, mivel az y-nak többféle jelentése is lehet. Kettős betűk második tagjaként (ny, ty, stb.) nem ejtjük külön, míg régies helyesírású és idegen szavakban i-ként (magánhangzó) vagy j-ként (mássalhangzó) ejtendő. Mivel egy ezeket kezelő szabály igen bonyolult, és semmiképpen nem lett volna teljes, ezért ezen eseteket kizárólag kivételszótárral kezeltük.

4. Szótagokra bontás

A szótaghatárokat a betűhatárok alapján állapítottuk meg. Ehhez a magyar nyelv szótagolási szabályát vettük alapul [Ferenc, 1992]. Eszerint minden magánhangzó külön szótagot alkot. A szótagot alkotó magánhangzó az ún. szótagmag. A szótag eleje a szótagmagot képező magánhangzót megelőző mássalhangzó (ha van ilyen), ellenkező esetben maga a szótagmagot képező magánhangzó. A szó elején lévő összes mássalhangzót a rá következő magánhangzó által alkotott szótaghoz rendeljük. A hosszú kettős betűket a magyar elválasztásnak megfelelően szótagoljuk. Eszerint a 'hosszú' és 'llyés' nevek 'hosz, szú' és 'ily, lyés'-ként lesznek szótagolva. A program nem valósít meg morfológiai elemzést. Az összetett, -ság, -ség képzős, igekötős, kvázi igekötős és leg-gel kezdődő szavak helyes elválasztását a kivételszótár tartalmazza.

Az elválasztási kivétel-szótárakban szereplő szavak elválasztását a szótárban előírt pontokon módosítja a program. Itt egyrészt több ezer, az általános szabályoktól eltérő elválasztású szó (például összetett szavak, bizonyos cz-s szavak, stb.), másrészt a magyar szabályokhoz képest eltérő elválasztású idegen eredetű cég és személynevek szerepelnek. Ezeket több ezer cég és rendhagyó személynevet tartalmazó adatbázisok elemzése, illetve az elemek meghallgatása révén alakítottuk ki.

5. Rövidítések és írásjelek feloldása

A kizárólag nagy betűket tartalmazó szavak többnyire cégnevek, amelyeket bizonyos esetekben betűnként (pl. IBM), máskor egybe (pl. MATÁV) kell felolvasni. Ezek kezelésére egy kivételszótárral kombinált szabályrendszert alkottunk. A szabályrendszer alapvetően a szó hosszát és a benne szereplő magánhangzók számát figyeli. A csupa nagybetűt tartalmazó rövidítések elé a rendszer a 'csupanagybetűvel' megjegyzést szúrja be (pl. az IBM 'csupanagybetűvel, í, bé, em'-ként kerül felolvasásra).

A felolvasandó nevekben szereplő rövidítéseket (pl. 'Kft', 'Bt', stb.) a rövidítéseket feloldó modul teljes alakjukra ('káefté', 'bété') cseréli. Hasonlóan a szövegben szereplő írásjelek is lecserélésre kerülnek.

6. A helyes leírást segítő megjegyzések beillesztése

A részletező modul egyik fontos feladata, hogy segítse az ügyfelet a hallott név helyes lejegyzésében. Ezért részletező megjegyzéseket is fűzünk a szótagoláshoz. Itt például a w-s, y-os, ly-os, kettős mássalhangzót, hosszú magánhangzót, stb., tartalmazó szótagok kerülnek kommentálásra. Így például a 'Way Kft.' 'vaj, duplavével, ipszilonnal, káefté'-ként lesz részletezve.

Az önmagukban álló árva betűkhöz helytelen lenne megjegyzés fűzése, ezért ezeket a következő modul egyszerűen a megfelelő szövegre cseréli. Így például egy önmagában álló 'w' 'duplavére' cserélődik.

A még nem lecserélt idegen betűk, tipikusan a q-betűk, egy nekik megfelelő helyettesítő szövegre cserélődnek. Például a 'Quarz' szó 'kú betű, u, arz'-ra cserélődik. Ez biztosítja, hogy a q-betűs helyesírás egyértelmű legyen a felhasználó számára. A q-betű kezelését, ugyanis nem lehetett más idegen betűk (pl. y, w, stb.) kezelésekor követett módszerhez hasonlóan megoldani.

A meghallgatásos tesztek során egyes szótagokban nehéz volt eldönteni, hogy azok hosszú, vagy rövid magánhangzót tartalmaznak-e. Ezért a rendszer hosszú magánhangzók esetén is kiegészítő megjegyzést szűr be (pl. az Ír Bt. 'ír, hosszúível, bété'-ként kerül felolvasásra).

7. Prozódiai címkék beillesztése

A szótagokra tagolt, részletezett szöveg felolvasása nem igényli bonyolult dalmenet alkalmazását. Meghallgatásos vizsgálataink alapján elegendőnek bizonyult a bemondás utolsó szótagjában levinni a hangmagasságot. A beszéd szintizátor dallammenetét a szövegbe helyezett speciális vezérlő jelzésekkel lehet vezérelni. A részletező modul egy ilyen speciális jelzést helyez az utolsónak kimondásra kerülő szótag elé.

A megfelelő ritmika kialakításának érdekében eltérő hosszúságú szüneteket iktattunk be a szótagok, illetve a hozzájuk tartozó megjegyzések közé. Meghallgatásos tesztek során több változat közül olyan értékeket választottunk, amelyekkel a legjobb volt a felolvasott szöveg érthetősége. Ezen szünetek is speciális jelzéseként (pl. [:pause 500]) kerülnek a szövegbe.

A meghallgatásos tesztek során arra is fényderült, hogy még hosszabb szünetek alkalmazása esetén is nehézséget okozhat a felolvasandó nevekben esetlegesen szereplő szóhatárok elkülönítése. Emiatt a szóhatár hollétének markáns jelölésére azt 'szóköz'-ként külön be is mondja a rendszer. Például a Balta Bán Kft 'bal, ta, szóköz, bán, káefté'-ként kerül bemondásra. Az ennek megfelelő szövegrészt a szünetekkel egyszerre illesztjük be.

A kimondott szöveg érthetőségét tovább növelendő, az egyes szótagokat lassabban, míg a hozzájuk fűzött megjegyzéseket gyorsabban olvassa a rendszert. Ezen sebesség-beállításokat jelző parancsoknak az elhelyezésére is itt kerül sor.

8. Összefoglalás

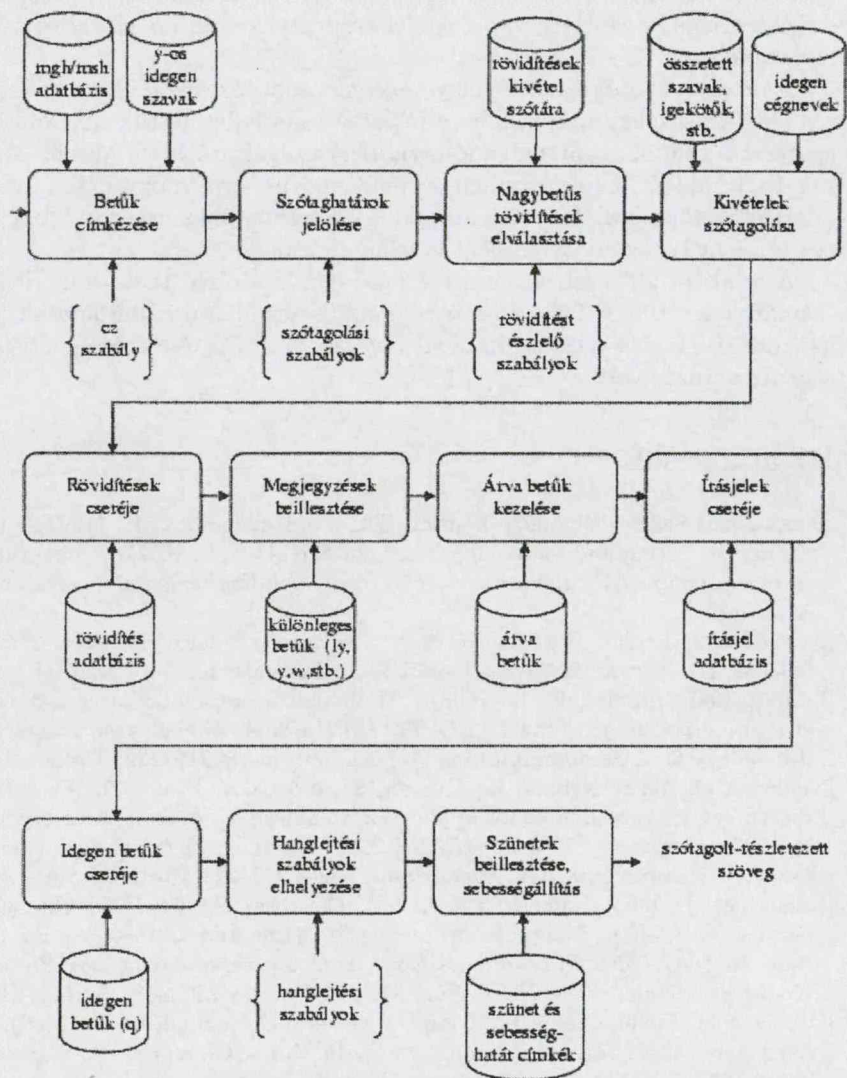
A fent leírt módon előkészített szöveget a név-és címfelolvasó szintetizátor modulja szótagolva, részletezve, megfelelő szünetekkel, intonációval és sebességgel olvassa fel.

A rendszert több-ezres személy és cégnév adatbázisokon végzett meghallgatásos tesztekkel ellenőriztük, illetve folyamatosan fejlesztettük. Az adatbázisokat egyrészt a velünk együttműködő távközlési szolgáltatótól (T-Mobile Magyarország Rt.) kaptuk. Ezek azonban személyiségi és egyéb jogvédelmi szempontok miatt érthetően korlátozottak voltak, így a teszteléshez más forrásból származó cég és személynév adatbázisokat is felhasználtunk.

A rendszer 2004 februárja óta érhető el a T-Mobile 1230-as ügyfélszolgálati számán keresztül. A 2004. szeptemberéig összegyűlt forgalmi adatok alapján a lekérdezések 63.5%-a egyszerű névfelolvasás, 17,5%-a részletező felolvasás, 19%-a pedig betűzés volt.

Hivatkozások

- [Bentley and Sedgewick, 1997] Bentley, J. L. and Sedgewick, R. (1997). Fast algorithms for sorting and searching strings. In *SODA: ACM-SIAM Symposium on Discrete Algorithms (A Conference on Theoretical and Experimental Analysis of Discrete Algorithms)*.
- [Ferenc, 1992] Ferenc, K., editor (1992). *Strukturális magyar nyelvtan, 2. kötet – Fonetika, 4.4. fejezet: Szótagolási szabályok*. Akadémiai Kiadó, Budapest.
- [Lundin, 1998] Lundin, F. J. (1998). The swedish automatic reverse directory service. In *Proceedings of the IVTTA'98, IEEE-ESCA Workshop on Interactive Voice Technology for Telecommunication Applications*, pages 219–222, Torino, Italy.
- [Nebbia et al., 1998] Nebbia, L., Quazza, S., and Salza, P. L. (1998). A specialized speech synthesis technique for application to automatic reverse directory service. In *Proceedings of the IVTTA'98, IEEE-ESCA Workshop on Interactive Voice Technology for Telecommunication Applications*, pages 223–228, Torino, Italy.
- [Németh et al., 2003] Németh, G., Zainkó, C., Kiss, G., Fék, M., Olasz, G., and Gordos, G. (2003). Language processing for name and address reading in Hungarian. In *2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2003)*, pages 238–243, Beijing, China.
- [Olasz et al., 2000] Olasz, G., Kiss, G., Németh, G., and Olasz, P. (2000). Profivox: a legkorszerűbb hazai beszéd szintetizátor. In Mária, G., editor, *Beszéd kutatás 2000*, pages 167–179. MTA Nyelvtudományi Intézete.
- [Spiegel, 1993] Spiegel, M. F. (1993). Coping with telephone directories that were never intended for synthesis applications. In *Proceedings of the ESCA-NATO/RSG 10 Tutorial and Workshop on Applications of Speech Technology*, pages 19–22, Lautrach, Germany.



1. ábra. A névfelolvasó részletező moduljának felépítése.