

Kulcsszó-előfordulások relevanciájának vizsgálata magyar nyelvű hangzó híryanagyokban¹

Gosztolya Gábor

MTA-SZTE Mesterséges Intelligencia Tanszéki Kutatócsoport,
6720 Szeged, Tisza Lajos krt. 103.
ggabor@inf.u-szeged.hu

Kivonat: Kulcsszókeresés során a feladat felhasználók által beírt kulcsszavak előfordulásainak megtalálása nagyméretű hangadatbázisokban. Egy adott kulcsszókeresési rendszer pontosságának meghatározásához ismernünk kell a kulcsszavak valós előfordulásait, mely feladatra léteznek automatikus módszerek, azonban az, hogy ezek eredményei mennyire esnek egybe az emberi elvárásokkal, nem egyértelmű. Ennek vizsgálatához néhány tesztalanyt kértünk meg, hogy azonosítsák a számukra releváns kulcsszó-előfordulásokat. Válaszaikat több szemszögből elemeztük: megvizsgáltuk, használatukkal mennyire változik meg kulcsszókereső rendszerünk pontossága; elemeztük, mennyire esnek egybe a válaszok egymással; valamint azt is megnéztük, hogy az egyes alanyok jellemzően milyen jellegű előfordulásokat tartottak relevánsnak.

1 Bevezetés

A kulcsszókeresési probléma (Spoken Term Detection, STD [6]) egy viszonylag új beszédtechnológiai terület, melyben a feladat különböző, felhasználó által bevitt kulcsszavak előfordulásainak megtalálása egy nagyméretű hangadatbázisban. Bár hasonló alapokra építkeznek, mint a beszédfelismerés, alapvető céljukban eltérnek: míg a beszédfelismerés változó bemondásokhoz meghatározni a pontos szöveges átiratot, jellemzően változatlan nyelvi és akusztikus modell mellett, kulcsszókeresésben a bemondások halmaza rögzített, míg a kulcsszavak változnak a felhasználás során.

Mint a mesterséges intelligenciabeli alkalmazások általában, egy kulcsszófelismerő rendszer is hangolható annak érdekében, hogy minél inkább az elvárásoknak megfelelően működjön. Ennek során egy rögzített felvételhalmazon és rögzített kulcsszókészletet használva értékeliük ki egy konfiguráció teljesítményét valamilyen pontosság-mértékkel, és ehhez hangoljuk az eljárás paramétereit. A kiértékeléshez azonban *annotált* hangfelvételekre van szükség: olyanokra, melyeknél előre meghatároztuk a kulcsszavak előfordulásainak pontos helyeit. Ez a feladat egyszerűnek tűnhet, amennyiben rendelkezésünkre áll a hanganyagok időzített szöveges átirata: ekkor azt te-

¹ Jelen kutatási eredmények megjelenését a „Telemedicina-fókuszú kutatások orvosi, matematikai és informatikai tudományterületeken” című, TÁMOP-4.2.2.A-11/1/KONV-2012-0073 számú projekt támogatja. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

kintjük egy kulcsszó tényleges előfordulásának, ahol a keresett szó teljes egészében, önállóan megtalálható. Ez azonban sokszor leegyszerűsítés, teljes mértékben figyelmen kívül hagyja például az összetett szavakat. Különösen így van ez ragozó nyelvek (mint amilyen a magyar is) esetében, ekkor ugyanis a kulcsszó toldalékolt alakjait is valós előfordulásnak kell tekintenünk, melyek automatikus meghatározása nem triviális.

A felvetett problémára a legjobb megoldás az lenne, ha valamilyen emberi címkézést használnánk, azonban ennek nyilvánvaló hátránya a nagy munkaigény, emiatt nagyobb adatbázisok felcímkézése elég drága. Az is várható, hogy egyes felhasználók véleménye egy-egy konkrét esetben eltér, ugyanakkor valamiféle „objektív” listára lenne szükségünk. Kérdéses, hogy az egyes felhasználók visszajelzéseit összegezve kaphatunk-e egy széles támogatottságú listát.

Jelen cikkben a kulcsszó-relevancia problémát vizsgáltuk, elsősorban a fenti szempontokra koncentrálva. Összeállítottunk egy kérdőívet egy magyar nyelvű hangadatbázis [1] kétséges kulcsszó-előfordulásairól, és felkértünk öt tesztalanyt, hogy ezek közül válasszák ki a szerintük relevánsnak tartott előfordulásokat; válaszaikat ezután több szempontból is elemeztük.

A cikk második fejezetében felvázoljuk a kulcsszókeresési feladatot és ismertetjük az abban széleskörűen alkalmazott pontossági metrikákat. A harmadik fejezetben részletesen leírjuk az alkalmazott automatikus kulcsszóelőfordulás-detektáló módszereket és a kérdőív összeállításának menetét. A negyedik fejezetben a felhasznált adatbázist és a kulcsszókereső rendszert ismertetjük; végül az ötödik fejezetben bemutatjuk az eredményül kapott pontossági értékeket, és részletesen elemezzük a különböző felhasználók válaszait.

2 A kulcsszókeresési feladat

A kulcsszókeresési feladatban felhasználók által beírt *kulcsszavak* előfordulásait keressük korábban rögzített (hang)felvételek egy halmazában. A kulcsszókereső rendszer előfordulás-hipotézisek listáját szolgáltatja, melyek mindegyike tartalmazza az előfordulás helyét (felvétel, kezdő és befejező időpontok), a kulcsszót és a hipotézis valószínűségét, mely szerint azok sorba rendezhetők. Más hasonló területekkel ellentétben a hipotézisek sorrendje nem lényeges, a valószínűség a hipotézisek szűrésére szolgál.

A hangfelvételek feldolgozása általában elég erőforrás-igényes, a felhasználó viszont joggal vár gyors választ, így tipikus a felvételek valamilyen mértékű előfeldolgozása; ez egy köztes reprezentációt eredményez, amelyben aztán a keresést végezzük. A több ismert reprezentáció közül jelen cikkünkben a legvalószínűbbnek talált fonémasorozatot használjuk, mely elég gyors keresést tesz lehetővé.

Cikkünk szempontjából persze a konkrét kulcsszókeresési algoritmus csak annyiban érdekes, hogy az általa visszaadott előfordulás-hipotézisek (melyek az egész cikkben változatlanok) hogyan illeszkednek a különféle módszerekkel meghatározott *releváns előfordulásokhoz*, és az utóbbiak hogyan befolyásolják a kulcsszókereső rendszer pontosságát. Ehhez azonban először definiálnunk kell a használt pontosság-metrikákat.

2.1 Az alkalmazott pontosságértékek

A kulcsszókeresési probléma egy információ-visszakeresési feladat, emiatt hagyományos IR metrikákkal: pontossággal (*precision*) és fedéssel (*recall*) is mérhető egy adott algoritmuskonfiguráció teljesítménye [6]. A legtöbb információ-visszakeresési területen a két metrikát azok (parametrikus) harmonikus közepével, az F-mértékkel (*F-measure*) szokás egyetlen értékke aggregálni, azonban a kulcsszókeresés területén más metrikák terjedtek el. Leggyakrabban a Figure-of-Merit (FOM) mérőszámot használják, mely az óránként és kulcsszavanként 1, 2, ... 10 hibás találat megengedése esetén elért fedési értékek számtani közepe. A másik elterjedt mérőszámot az amerikai National Institute of Standards and Technology (NIST) vezette be 2006-os kulcsszókeresési versenyén: ez az aktuális kulcsszó-súlyozott érték (*Actual Term-Weighted Value*, ATWV), mely a következőképpen definiált:

$$\text{ATWV} = 1 - \frac{1}{T} \sum_{i=1}^T (P_{\text{Miss}}(t) + \beta P_{\text{FA}}(t)), \quad (1)$$

ahol $P_{\text{Miss}}(t)$ az adott kulcsszó eltévesztésének, $P_{\text{FA}}(t)$ pedig hibás találatának valószínűsége; azaz

$$P_{\text{Miss}}(t) = 1 - \frac{N_{\text{corr}}(t)}{N_{\text{true}}(t)} \quad \text{és} \quad P_{\text{FA}}(t) = 1 - \frac{N_{\text{FA}}(t)}{T_{\text{speech}} - N_{\text{true}}(t)}, \quad (2)$$

ahol $N_{\text{corr}}(t)$ az adott kulcsszó helyes találatainak, $N_{\text{true}}(t)$ a tényleges előfordulásainak, $N_{\text{FA}}(t)$ a hamis találatainak száma, T_{speech} pedig az átfésülendő felvételek összhossza másodpercben mérve [3]. β értéke általában 1000. Egy, a használt annotációval tökéletes összhangban működő rendszer ATWV pontszáma 1,0, egy olyané, amely egyáltalán nem ad vissza találatokat, 0,0. Feltételezve, hogy T_{speech} lényegesen nagyobb, mint $N_{\text{true}}(t)$, egy olyan rendszer, amely az összes elvárt előfordulást megtalálja, de minden kifejezésre óránként 3,6 hamis találatot produkál, szintén 0,0 értéket fog kapni, így ez a metrika jóval szigorúbb, mint a FOM. További különbség, hogy az ATWV az összes visszaadott találatot figyelembe veszi, míg FOM esetén csak a valószínűbbeket. Kísérleteink során mindkét metrikát alkalmaztuk.

3 A releváns előfordulások meghatározásának módjai

A következőkben azt ismertetjük, milyen stratégiákat alkalmaztunk, hogy meghatározzuk a kulcsszavak előfordulásainak helyeit a hangfelvételek szöveges átirata alapján.

3.1 Automatikus módszer

A legkézenfekvőbb megoldás (elvárjuk a kulcsszó előfordulását önálló szóként, ill. szószorozatként) a bevezetőben már említett okok (toldalékolás, összetett szavak) miatt nem alkalmazható, azonban annak egy módosított változata már igen: ekkor azt várjuk el, hogy a kulcsszó az átiratban teljes egészében bukkanjon fel egy szóban. Ezzel a ragozott szóalakokat is elfogadjuk. A magyar nyelv ragozási szabályait figyelembe véve a magánhangzóra végződő kulcsszavak esetében a hosszúra váltó magánhangzós változatot is elfogadtuk (pl. *Amerika – Amerikában*). Persze ez a megoldás sem tökéletes, különösen rövid kulcsszavakra jellemző, hogy sokszor fordulnak elő más szó belsejében, így sok téves riasztáshoz vezetve.

1. táblázat:

Relevánsnak minősített előfordulások száma a különböző alkalmazott módszerekkel a validációs és teszt adatbázisrészeken

Módszer	Validációs	Teszt
Automatikus	381	709
1. alany	365	690
2. alany	368	689
3. alany	396	732
4. alany	366	699
5. alany	367	697
Alanyok (többségi szavazás)	367	697
Egyértelmű	334	651

3.2 Emberi annotálás

A másik lehetőség, hogy akkor tekintünk egy előfordulást relevánsnak, amennyiben egy ember annak tekinti. Bár nyilván ez a legpontosabb módszer, hiszen pontosan akkor lesz relevánsnak minősítve egy előfordulás, amennyiben egy ember úgy gondolja, hogy az valóban releváns; nagyobb archívumok emberi annotálása azonban elég drága. Jelen cikkünkben viszont éppen arra voltunk kíváncsiak, hogy milyen változásokat okoz az emberi vélemények figyelembe vétele, és a felvétel-adatbázis sem volt túl nagynak mondható, így kísérleteinkben alkalmazhattuk ezt a megközelítést.

El szeretnénk volna kerülni, hogy az alanyok a többórányi hangfelvétel teljes leiratát annotálják az összes, a tesztjeinkben szereplő kulcsszóra, így automatikus módszerekkel leszűkítettük a lehetséges releváns előfordulások halmazát, és egy kérdőívre gyűjtöttük őket. Betűalapú illesztési távolságot használva megkerestük azokat a helyeket, ahol a kulcsszavakhoz hasonló betűsorozatok fordultak elő; legfeljebb az adott kulcsszó hosszának 30%-át kitevő betűbeszúrás, -törlés és -cserét engedtünk meg (tehát egy 10 betűvel leírható kulcsszó esetén legfeljebb három művelettel elő kellett tudni állítani azt). Mivel a lista még így is túl hosszú volt, azokat a potenciális előfor-

dulásokat automatikusan relevánsnak tekintettük, melyeknél *szó elején és teljes egészében* fordult elő az adott kulcsszó. (Ezekre az 5. fejezetben *egyértelmű* releváns előfordulásként fogunk hivatkozni.) Mindezt abból a megfontolásból tettük, hogy ezek nagy eséllyel a kulcsszó ragozott alakjai, és habár ez nem minden esetben teljesült (pl. bizonyos összetett szavaknál), összességében elég jó közelítésnek találtuk, és hatékonyan csökkentette a kérdőív hosszát.

Végül ezt a kérdőívet töltöttük ki öt tesztalannyal; az általuk megjelölt előfordulások és az egyértelmű előfordulások halmazának unióját tekintettük szerintük releváns előfordulásoknak.

Az 1. táblázat mutatja a különböző automatikus módszerek és az egyes alanyok által relevánsnak minősített előfordulások számát. A kérdőív 111, illetve 242 potenciális előfordulást tartalmazott (a fejlesztési és a tesztelési halmazokra vonatkoztatva), melyek közül az alanyok 31-62-t, illetve 38-81-et választottak ki. A számok azt is tükrözik, hogy az alanyok (a 3. alany kivételével) alapvetően hasonlóan ítélték meg a potenciális releváns előfordulásokat (bár ehhez a kulcsszókeresési rendszer pontosságértékeit is érdemes megvizsgálni), és gyökeresen különböző módon, mint a két alkalmazott automatikus módszer. Mivel arra is kíváncsiak voltunk, hogy elérhető-e valamiféle konszenzus az alanyok között, a táblázatokban feltüntettük az egyszerű többségi szavazáshoz tartozó értékeket is.

4 Technikai megoldások

Mielőtt bemutatnánk és elemeznénk a teszteredményeket, még be kell mutatnunk, hogyan párosítjuk össze az előfordulás-hipotéziseket a releváns előfordulásokkal, valamint ismertetnünk az alkalmazott kulcsszókeresési rendszert és az adatbázist.

4.1 Az előfordulás-hipotézisek és a releváns előfordulások összepárosítása

Az irodalomban több megoldást is találunk a kulcsszóhipotézisek és -előfordulások összepárosítására. Természetesen a hipotézisnek és a tényleges előfordulásnak ugyanabban a felvételben kell lennie, és ugyanahhoz a kulcsszóhoz kell tartoznia. Mindezekon túl azt is elvárjuk, hogy a hipotézis ugyanabban az időpontban hangozzon el, mint a tényleges előfordulás, azonban ezen nyilvánvalóan nem érthetjük azt, hogy a kezdő- és végpontok is tökéletesen egybeessenek. Elvárhatjuk például, hogy ezek valamilyen határon belül legyenek; [3] esetében a tényleges előfordulás közepétől legfeljebb fél másodpercre kell esnie a hipotézisnek, míg [7] akkor párosítja össze a hipotézist egy tényleges előfordulással, amennyiben a hozzájuk tartozó időintervallumok metszik egymást. Mi az utolsó megoldást alkalmaztuk, részben tekintettel a magyar nyelv ragozó voltára, mely eléggé megnehezíti a szigorúan vett kulcsszó pontos kezdő és befejező időpontjainak meghatározását.

4.2 A kulcsszókeresési rendszer

Kísérleteinkben saját kulcsszókeresési keretrendszerünket használtuk (részletesen lásd [2]). Ebben a hangfelvételeken először beszédfelismerési lépéseket végzünk, jelen esetben egy nagypontosságú, kétmenetes neuronhálós fonémaosztályozási módszerrel alkalmazva [5]. Az eredményül kapott fonémasorozatot letároljuk, és erre illesztjük a beírt kulcsszó fonetikus átíratát. Az illeszkedés mértékét illesztési távolság (*edit distance*) metrikával mérjük, fonémánként eltérő műveleti költségeket használva, melyeket a fonémaosztályozó tévesztési mátrixából számítunk [4].

2. táblázat:

Kulcsszófelismerési pontosságok alakulása a különböző alkalmazott módszerek függvényében

Módszer	FOM	ATWV
Automatikus	88,72%	56,84%
1. alany	88,35%	52,32%
2. alany	87,39%	48,00%
3. alany	88,85%	60,23%
4. alany	88,15%	52,90%
5. alany	88,22%	53,05%
Alanyok (átlag)	88,19%	53,30%
Alanyok (medián)	88,22%	52,90%
Alanyok (többségi szavazás)	88,22%	53,07%
Egyértelmű	87,94%	44,77%

4.3 A felhasznált adatbázis

A kísérletekhez 70 híradót rögzítettünk nyolc tévécsatornáról (ATV, Hálózat TV, Hír TV, M1, M2, Rtl, Tv2) [1]. A felvételeket néhány mondatos blokkokra vágtuk; közülük jelen cikkben csak azokat használtuk, melyekben szépen artikulált beszéd hallható és a háttérzaj minimális. A 70 híradót 44-9-17 arányban osztottuk fel tanítási, fejlesztési és tesztelő blokkokra (időtartamot tekintve ez kb. 5 és fél óra – 1 óra – 2 óra), ügyelve arra, hogy a tévécsatornák mindegyikéből kerüljön mindegyik részhalmozba. A felvételek mindegyikét legépeltük, az ortografikus átíratot utólag is ellenőriztük. Az alkalmazott 50 kulcsszót a felvételekben gyakran előforduló főnevek közül választottuk ki; illeszkedve a felhasználói igényekhez, jelentős részük (18 darab) tulajdonnév volt. Hosszuk 6-16 fonéma, 2-6 szótag között alakult.

5 Eredmények

5.1 Kulcsszófelismerési pontosságok

A 2. táblázat tartalmazza az elért pontosságokat a különböző, a releváns kulcsszó-előfordulásokat detektáló módszerek esetén. Látható, hogy a FOM értékek gyakorlatilag változatlanok, míg az ATWV pontosságok elég nagy skálán (48,00%-tól 60,23%-ig) mozognak. A tesztalanyokhoz tartozó pontosságok nagyban eltérnek a két (*automatikus*, illetve *egyértelmű*) automatikus módszerhez tartozóktól is: ez alapján a felhasználói elvárásokhoz képest az egyik automatikus módszer jellemzően túl megengedő, a másik pedig túl szigorú. A többségi szavazással elért pontosságérték (53,07%) nagyon közel áll három tesztalanyéhoz (1., 4. és 5.), valamint az átlagos és a medián pontosságértékhez is. Ez azt jelzi, hogy egyszerű többségi szavazással valószínűleg elérhető egy, a gyakorlatban jól teljesítő konszenzusos előfordulás-lista.

5.2 A felhasználói válaszok elemzése

A pontosságértékek változásainál is érdekesebb kérdés, hogy az egyes előfordulásokat hogyan értékelték az egyes alanyok, és a vélemények mennyire hozhatók közös nevezőre. A következőkben ezeket a konkrét eseteket fogjuk körüljárni.

A csak korlátozott nyelvi információt hasznosító kulcsszókereső megközelítések ismert hátránya, hogy hajlamosak az (általában rövid) kulcsszavakat más szavak belsejében is „megtalálni”, és így sok hamis riasztást generálni. Esetünkben ez a *kormány* kulcsszóval fordult elő jelentősebb számban, mely valóban megtalálható az *önkormányzat* szó belsejében, így ezeket az előfordulásokat az automatikus keresőmódszer is relevánsnak minősítette; ugyanakkor az öt alanyból négy vélte úgy, hogy ezek hamis riasztások. Kulcsszókereső rendszerünk, mely csak az akusztikus információra támaszkodhatott, természetesen szintén megtalálta ezeket az „előfordulásokat”.

Az automatikus módszerben megengedtük, hogy a kulcsszó szóvégi magánhangzója hosszúra váltsón (a többi magánhangzó viszont nem). A *vasút* kulcsszó esetében hasonló dolog történt, csak ellenkező előjellel: mindegyik alanyunk úgy vélte, hogy a *vasutas* szó is a *vasút* kulcsszó releváns előfordulása. Ugyanakkor, habár hangtanilag tökéletesen ugyanez az eset a *miniszter* kulcsszó és a *minisztérium* szó, a megkérdezett alanyok közül mégis mindössze egy sorolta ezt a releváns előfordulások közé.

További nagy csoport volt a kulcsszavak között bizonyos személyneveké: *Angela Merkel* (német kancellár), *Bajnai Gordon* vagy *Orbán Viktor* (magyar miniszterelnökök). Kulcsszóként a teljes név volt megadva, időnként azonban a felvételekben csak vezetékneveikkel hivatkoztak rájuk. Az összes alany egyetértett azzal, hogy ezek is releváns előfordulások, bár csak a keresett kulcsszavak fele fordult elő. Megjegyzendő, hogy mivel illesztési távolságot használva állítottuk össze a kérdőívet, azon csupán azok az előfordulások szerepelhettek, ahol a szövegekörnyezet a hiányzó keresztnévhez igen hasonló volt (pl. „amely Merkel”, „Bajnai-kormány”, „Orbán-kormány”).

Ehhez igen hasonló eset volt a *rendőrség* kulcsszóé: többször is szerepelt a kérdőívben a *rendőr* szó, melyet az ötből három alany tartott releváns előfordulásnak annak ellenére, hogy itt a kulcsszó tartalmazta a ténylegesen előforduló szót. Ez feltehetőleg azt tükrözi, hogy ezen alanyok számára a két fogalom szorosan összekapcsolódik. Hasonló viszonyt jelez a *gázár* kulcsszó esete is: a többször is szereplő „gáz ára” szókapcsolatot ugyanis az összes alany a kulcsszó releváns előfordulásának tekintette.

A fenti példák esetében az alanyok általában egyetértettek egymással, a válaszokat azonban nemigen lehetne automatikusan megjósolni. Ha egy elhangzott szó teljes egészében tartalmazza a keresett kulcsszót, az általában releváns előfordulás; bizonyos esetekben (*kormány*) ugyanakkor nem az, máskor pedig a kulcsszó tartalmazza a ténylegesen elhangzott szót (*rendőrség*). A kulcsszó szóvégi magánhangzója hosszúra válthat, és ez időnként más magánhangzókkal is előfordulhat (*vasút*), más esetekben viszont nem (*miniszter*). A *gázár* kulcsszó esete valószínűleg egyáltalán nem kezelhető automatikusan: amennyiben kulcsszavakon belül akárhol engedélyezünk szóhatárokat, az rengeteg hamis riasztáshoz vezethet. Viszont ha ismertebb személyeket keresünk, célszerű a kulcsszót csak a vezetéknevek választani (*Merkel, Bajnai, Orbán*).

Amikor a megkérdezett alanyok egy-egy hipotézis besorolásakor nem értettek egyet, szinte mindig négy az egyhez aránylottak a szavazatok; összesen négy helyen alakult ez három a kettőhöz. Ez azt sugallja, hogy szinte minden esetben elérhető egy elfogadott konszenzus, azaz létrehozható olyan címkézés, mely szinte teljesen egybeesik az emberek által elvárt viselkedéssel. (Ezt természetesen érdemes lenne ötnél lényegesen több alanyra is megvizsgálni.) Ezt kulcsszókeresési rendszerünk pontosságértékei is alátámasztották: amennyiben szavazásnál azt vártuk el, hogy legalább négy alany értsen egyet az adott előfordulás megítélésében, a pontosságértékek alig változtak, egyhangú eredmény elvárása esetén viszont számottevően csökkentek.

Az emberi annotálással elért pontszámokat az automatikus módszerekéihez hasonlítva egyértelmű, hogy alapvetően különböznek: mikor csak a tiszta előfordulásokat tekintettük relevánsnak, az ATWV értéke 44,77% lett, mely a többi előforduló pontosságértékhez mérten alacsony (valószínűleg a sok hamis riasztás miatt); mikor viszont a standard automatikus módszert alkalmaztuk, az túl megengedőnek bizonyult, amely az előálló, irreálisan magas 56,84%-os ATWV értékben is tükröződik.

6 Konklúzió

Jelen cikkünkben szokatlan nézőpontból vizsgáltuk meg a kulcsszókeresési problémát: azt elemeztük, hogy az automatikusan előállított kulcsszó-előfordulások mennyire egyeznek a felhasználói igényekkel. Ehhez tesztalanyokat kértünk meg, hogy jelöljék meg, mely potenciális előfordulásokat tekintik valóban relevánsnak. A válaszokat elemezve azt találtuk, hogy, habár nem volt két pontosan ugyanúgy válaszoló alany, összességében a válaszok egymáshoz nagyon hasonlóan bizonyultak, és egyszerű többségi szavazással egyértelmű konszenzus volt elérhető. A kipróbált automatikus eljárások azonban vagy túl optimisták, vagy túl pesszimisták voltak, és a tesztalanyok válaszait részletesen megvizsgálva azt sem tartjuk valószínűnek, hogy automatikus (szintaktikai) eljárásokkal azok reprodukálhatóak lennének.

Hivatkozások

1. Gosztolya, G., Tóth, L.: Kulcsszókeresési kísérletek hangzó híryanagyokon beszédhang alapú felismerési technikákkal, Proc. MSZNY (2010) 224–235
2. Gosztolya, G., Tóth, L.: Spoken Term Detection Based on the Most Probable Phoneme Sequence, Proc. SAMI (2011) 101–106
3. NIST: The Spoken Term Detection (STD) Evaluation Plan, National Institute of Standards and Technology (NIST), Gaithersburg, USA, <http://www.nist.org/speech/tests/std> (2006)
4. Szöke, I., Schwarz, P., Matejka, P., Karafiát, M.: Comparison of Keyword Spotting Approaches for Informal Continuous Speech, Proc. Interspeech (2005)
5. Tóth, L.: A Hierarchical, Context-Dependent Neural Network Architecture for Improved Phone Recognition, Proc. ICASSP (2011) 5040–5043
6. Wang, D.: Out-of-Vocabulary Spoken Term Detection, PhD thesis, Univ. Edinburgh (2010)
7. Young, S.J. et al: The HMM Toolkit (HTK) (software and manual), <http://htk.eng.cam.ac.uk/> (1995)