Magyar Ispell helyesírás-ellenőrző

dokumentáció

Németh László
Szofi Magyar–Amerikai Informatikai Oktatóközpont

2002. június 21.

Tartalomjegyzék

1. fejezet

Áttekintés

1.1. Bevezetés

A Magyar Ispell szabadszoftver projekt célja egy professzionális szintű magyar helyesírási szótármodul, illetve az ehhez kapcsolódó helyesírás-ellenőrző, és -javító program fejlesztése mind az irodai, mind a tudományos élet számára.

A projekt eredményes voltát bizonyítja, hogy a munkát jelenleg is már több ezren használják.

Az ingyenes, és a Magyar Szabadszoftver Alapítvány szervezésében honosított OpenOffice.org programcsomagnak szerves része a Magyar Ispell szótármodul, és a Magyar MySpell helyesírás-ellenőrző.

Az OpenOffice.org egy professzionális, és a Sun Microsystems termékeként is kapható MS Office-kompatibilis irodai csomag Linux és Windows operációs rendszerekre. (Mivel a windowsos és a linuxos magyar OpenOffice.org 1.0 változat 2002 júliusában jelent/jelenik meg nagy példányszámú számítástechnikai lapok lemezmellékletén, várhatóan a Magyar Ispell/MySpell felhasználók száma ugrásszerűen növekedni fog.)

A PrímPosta, az egyik legnagyobb magyar webmail-szolgáltató a közelmúltban átállt a Magyar Ispell/MySpell használatára a MS Office-ból jól ismert Helyes-e?-ről, lehetővé téve több ezer felhasználó leveleinek a korábbinál lényegesen egyszerűbb javítását.

A TEX/LATEX szedőrendszer használói, mint pl. egyetemi hallgatók, tudományos szakemberek, vagy a tudományos művek szedésével és kiadásával foglalkozó TypoTEX kiadó is, a Magyar Ispell potenciális felhasználói közé tartoznak.

Végül, de nem utolsó sorban a nemzetközi Linux terjesztéseknek, mint a SuSE Linux és a Debian GNU/Linux, már része a magyar Ispell szótármodul.

1.2. Tulajdonságok

A Magyar Ispell és MySpell rendszer alapvető tulajdonságait a következő felsorolás foglalja össze. A felsorolás a Morphologic cég Helyes-e? Microsoft Word 2000-es változatával való összehasonlítás is egyben. Elöljáróban le kell szögezni, hogy mélyreható vizsgálatokat a Helyes-e? kapcsán csak a *j/ly* betűs szavak esetében sikerült végezni.

Magyar Ispell és MySpell	Helyes-e?

1.2.1. Licenc

GNU GPL szabadszoftver licenc: a szótár és a helyesírás-ellenőrző program ingyenesen hozzáférhető és használható akár kereskedelmi célra is. Szabadon módosítható és terjeszthető. Más GPL-es programokba (pl. OpenOffice.org) beépíthető.	Nem ingyenes, illetve a Széchenyi-terv támogatásával Linuxra lefordított változat csak otthoni, nonprofit cél esetében használható majd ingyenesen. A forráskód zárt. Nem terjeszthető. Más rendszerekbe licencdíjért beépíthető (pl. Magyar Office).

1.2.2. Elvi alapok

Hasítótáblában tárolt tőszavak. Ragozási szabályok, és osztályok hatékony illesztéssel. A felismerés időigénye tőszavak esetében konstans, ragozott szavak esetében az időigény felső határa a ragozási osztályok számának logaritmusával arányos. Az algoritmus könnyen bővíthető.	Véges állapotú automata. A felismerés időigénye ragozott szavak esetében is egyenesen arányos a szó hosszával. Kis méretű szótárállományok.

1.2.3. Szókincs

Az alapszókincset toldalékolt alakjaival jól lefedő 10 000 ragozási szabály és 100 000 szóból álló, folyamatosan bővülő szókincs, szaknyelvi modulokkal (jelenleg a matematika, és a magyar helységnévtár modul a legteljesebb). A legtipikusabb tévesztések (j/ly, i/í) vonatkozásában a szóanyag ellenőrzött.	Az alapszókincset jól lefedő szótár. A problémás szavak egy részénél elfogadja a szó mindkét változatát (kompatíbilis/kompatibilis, pozícionálás/pozicionálás, húgyos/hugyos stb.).

1.2.4. Bővíthetőség

A szótár felépítése egyszerű és forrásszinten is könnyen bővíthető, akár új szaknyelvi modulok kialakításával is (minden ilyen irányú szándékot szívesen veszünk, és támogatunk). Lehetőség van a futásidejű bővítésre is, sőt egy a szótárban már meglévő szó megadásával a (jelenleg még csak az alkalmazás futásának idejére) felvett saját szavainkat is képes továbbragozni a Magyar MySpell.

Saját szótárakba felvehetünk új szavakat. Új szaknyelvi modulok létrehozására nincs közvetlen lehetőség. A felvett szavak továbbragozása mintha lehetséges lenne az MS Office egyes változataiban, de rövid keresés után itt most nem sikerült megtalálni ezt a funkciót.

1.2.5. Javítás

A Magyar MySpell program kifinomult helyesírás-javítási képességekkel rendelkezik.

Szegényes javaslattevés.

Javítja az összes lehetséges típusú – betűkihagyás, felesleges betű, szomszédos betűk cseréje vagy betűtévesztés hatására előállt – hibát.

Az egy betű távolságra lévő hibák egy részét javítja csak.

Javítja az egynél több karaktert érintő nyelvfüggő tipikus hibákat; jelenleg mintegy harminc mássalhangzó-alkalmazkodásból, nyelvjárási különbségekből stb. fakadó hiba esetében (pl. báttyát→bátyját, hüttő→hűtő, dijját→díját, kódexxel→kódexszel, stb). A javítási cseretáblázat akár a felhasználó által is tetszőlegesen bővíthető.

Egy-két gyakoribb több betűtávolságra lévő hibát javít (j/ly, ly/j, ggy/gyj, gyj/ggy, nny/nyj, de pl. a tty/tyj (báttya/bátyja) tévesztést már nem ismeri. A lista nem bővíthető.

Több javítási javaslat esetén a valószínűbbek előre kerülnek (pl. i/í, o/ó, u/ú, ü/ű tévesztések, a javítási cseretáblázattal orvosolható hibák, végül a betűk gyakorisága alapján).

Több javaslat esetén a sorrend esetleges.

A szótárba külön fel nem vett összetett szóösszetételek esetében is javaslatot tesz (pl, macskabukenc→macskabukfenc, papagálytoll→papagájtoll)

A szótárba külön fel nem vett összetett szavak esetében nem tesz javaslatot.

A formailag helyes, de értelmetlen és megtévesztő szóösszetételek kiszűrése (pl. szervíz, elitélt, ünnepéjes, vizitorma, vasárú). Ez korábban az egyes szavak szóösszetételben való megjelenésének letiltásával valósult meg, ami nem a legszerencsésebb (mivel teljes szóbokrok elfogadását korlátoztuk ezzel). A Magyar MySpell jelenlegi változata az összes olyan szóösszetételt hibásnak jelzi, ami előállhat egy helyes szóalak jellemző hibás leírásából.

A formailag helyes, de értelmetlen és megtévesztő szóösszetételek (egy részének) elfogadása (pl. szervíz, elitélt, ünnepéjes helyes szónak van elfogadva).

A Magyar MySpell különösen e-mailek javításánál hasznos képessége, hogy az ékezetek nélkül írt szavakra is helyes javaslatot tesz.

1.2.6. Ragozás

A köznapi szövegek szempontjából kielégítő. A ragozási hiányosságok köre folyamatosan szűkül és jól körülhatárolható. Jelenleg a következő tulajdonságok okozzák a problémákat:

A köznapi szövegek szempontjából kielégítő. Egy-két régebbről ismert hiányosság az új változatokban is megtalálható.

A legtöbb hibát a hiányos és/vagy hibás besorolások adják. A Magyar Ispell a szókincsbővítés fázisában van, és egy-két nagyobb szólista csak részben feldolgozott. Godó Ferenc 10 000 fel nem ismert szóalakot tartalmazó második listájának több, mint 10%-a az ingadozó -ja/-je -a/-e ragozással kapcsolatos (a lista nagy részét az új tőszavak teszik ki).

A -szerű utótag esetében nem fogadja el a Helyes-e? az egyszerűsítő írásmódot: mésszerű, gipsszerű, stb. hibás (MHSz §94). Elfogadja viszont a hibás alakokat: ésszerű, mészszerű, gipsszerű, sőt dzsesszszerű (helyesen: dzsessz-szerű).

Hangkivetős[1] és nyitótövek[2] esetén a melléknévképzős alakok feldolgozása még hiányzik a 3-6 szabályhoz szükséges szótagszám helyes megállapításához.

Hiányzó -ék jel (Kovácséknál).

[1] Lásd alant 18. l. és pl. [3, 804 skk.].

[2] Nyitótövek azok, melyekhez toldalék előtt a vártnál nyíltabb magánhangzó járul, pl. had, fül, lásd pl. [3, 923 skk.].

Igeképzők nincsenek alkalmazva minden főnévhez automatikusan (-z(ik), -ít/-sít, -kodik/-kedik/-ködik), hanem a gyakori képzett alakok vannak külön fölvéve. Az említett fel nem ismert szavakat tartalmazó listában a szavak 1%-a ilyen.	Képzőkkel kapcsolatos problémák: pl. a `következőek` szóra a javaslat a `követezőek`. Itt a folyamatos melléknévi igenév nem kapja meg a melléknév ragját, és a -zik képző általánossá tétele miatt növekszik a valószínűsége, hogy egy elütéssel a hibás alakok – itt a `következőek` helyett írt `követezőek` – is elfogadásra kerüljenek.
Az általános -gat, -get gyakorító képző szintén hiányzik, ami a Magyar Ispell következő változatában már elképzelhető, hogy megjelenik. (Az említett listában 15 ilyen szó.)	A feddd alak elfogadása.

1.2.7. Egybeírás

Az esetek nagy részében kielégítő. A kötőjeles különírás ellenőrzése fejlesztésre szorul.	Az esetek nagy részében kielégítő.
Hangkivetős és nyitótövek esetén a melléknévképzős alakok feldolgozása még hiányzik a 3-6 szabályhoz szükséges szótagszám helyes megállapításához (kelkáposztaleveles).	Négy tagból álló, de helyes szóösszetételek hibásnak jelzése (pl. barnakőszénkoksz, diófalevélszár).
Hiányzik bizonyos alakok előfordulási helyének az összetett szó elejére (mint pl. számok a mennyiségjelzős összetételeknél, pl. kutyaöt), illetve végére való korlátozása (hangkivetős, és nyitótövek ragozott alakjai, pl. pockokcsapda).	Nagy szótagszámú, de helyes szóösszetétel hibásnak jelzése (peronoszpóramegbetegedés).
A -szerű mintájára a -féle, -fajta, -nemű, -beli is képzőszerűek, és nem növelik a szavak számát a 3-6 szabály alkalmazásánál.	A *kis*, és a *nagy* melléknevek (előtagként) és a befejezett melléknévi igenevek (utótagként) szerepelhetnek korlátozások nélkül a szóösszetételekben pl. (nagy+meglepetéssel, elit+élt). Ugyanígy viselkedik az *ily* és az *oly* szó is.

1.2.8. Ly/j

A Helyes-e?-vel való összevetés alapján hibátlan az ly-t illetve j-t tartalmazó szavak kezelése tekintetében.

Hibás alakok (és ezek ragozott változatai): csevely, estéj, karvaj, kordéj, rostéj, szegéj, szeméj, szentéj, szeszéj, tartáj, tökéj, ünnepéj, olytat, súlytat, bolytorján, kardbolyt, tojássárgálya, vállbolyt, bolyár, bolytár, súlytó, restilye, ilyesztendő, ugorlyuk, csuklyalak, stb. Ezek között egy konkrét tévesztés van (a csevely), a többi (néha egészen hibás) szóösszetételként áll elő.

1.3. Összefoglalás

A két helyesírás-ellenőrző mindegyike tartalmaz még javítani valót, mind a szókincset, mind a ragozási szabályokat tekintve.

A Helyes-e? 2000 a közel tízéves fejlesztőmunka, jelentős állami támogatás és kereskedelmi szoftver volta ellenére súlyos hibákat tartalmaz (pl. a -szerű toldalék esetében az egyszerűsítő írásmód hiányát és az ly/j tévesztéseket). A Helyes-e? javítási képességei kezdetlegesek, sem az elütéseket, sem a tipikus helyesírási hibákat nem képes megfelelő módon javítani.

A Magyar Ispell/MySpell kiemelhető kiváló javítási képességei és ingyenessége, valamint bővíthetősége (szabadszoftver volta és jól kidolgozott keretrendszere illetve forráskódja) miatt. Ez és egyéb képességei lehetővé teszik, hogy akár az Akadémiai Helyesírási Bizottság által régóta kívánatosnak tartott szaknyelvi helyesírás-ellenőrző alapjául szolgáljon.

2. fejezet

Fejlesztői dokumentáció

2.1. Helyesírás-ellenőrzők ismertetése

Számos helyesírás-ellenőrzésre szolgáló program, illetve függvénykönyvtár áll a szabadszoftver-fejlesztők rendelkezésére.

2.1.1. Aspell

Angol nyelvterületen a legkedveltebb helyesírás-ellenőrző szabadszoftver az Aspell, Kevin Atkinson munkája. Az Aspell az angol nyelv fonetikai információinak ismeretében egy speciális algoritmussal képes nagymértékű alaki eltérések esetén is helyes javaslatot tenni.

Sajnos a programból hiányzik a nagy mennyiségű ragozási szabály hatékonyan kezelésének képessége, ezért nem megfelelő a magyar és sok egyéb nyelv számára sem.

2.1.2. Ispell

Az Ispell egyenes ági leszármazottja R. E. Gorin 1971-ben, PDP-10-re assembly-ben írt Spell programjának. A Spell-t Pace Williamson írta át C nyelvre 1983-ban. Geoff Kuenning 1987 és 1988 között nagyobb változtatásokat eszközölt a kódon, amelynek számunkra legjelentősebb lépése az affixum tömörítési technika bevezetése volt, az affixum tömörítéssel ugyanis lehetővé vált nagyszámú ragozási szabály futásidejű hatékony kezelése is. Gyakorlatilag ennek köszönhető, hogy az Ispell-hez több tucat nyelv helyesírási szótármodulja készülhetett el, köztük olyan agglutináló nyelveké, mint a finn és a magyar.

A magyar nyelv vonatkozásában az is kiemelhető, hogy az Ispell szabályozott szóösszetételképzési lehetőséggel rendelkezik, vagyis a szóösszetételképzést korlátozhatjuk a szavak egy bizonyos részhalmazára.

Az Ispell egyéb kiemelhető tulajdonságai:

- De facto szabvány a szabadszoftverek világában. A Free Software Foundation GNU pro-

jektjének alapértelmezett helyesírás-ellenőrzője, és a legtöbb Linux terjesztésben is megtalálható.

- A csőben használható felületének köszönhetően számos programba „beágyazható” (pl. Emacs, Kword, LyX).

2.1.3. Pspell

A Pspell függvénykönyvtár Kevin Atkinson kezdeményezése, ami az Aspell és az Ispell előnyeit hivatott a forráskód egybefogásával biztosítani.

2.1.4. MySpell

A MySpell az OpenOffice.org irodai programcsomaghoz készült BSD licencű helyesírás-ellenőrző függvénykönyvtár.

2001-ben kezdte fejleszteni Kevin Hendricks azzal a szándékkal, hogy a GPL licenc alatt is elérhetővé tett, de a licencelt szoftverkomponensektől (mint a nyelvi modulok) is megszabadított StarOffice – új nevén OpenOffice.org – helyesírás-ellenőrzővel rendelkezhessék.

A MySpell egy kicsi, C++-ban megvalósított program, amely az Ispell tömörítési algoritmusán alapszik és kis változtatással képes az Ispell szótármodulok használatára is.

A MySpell-nek az Ispell-lel szemben több határozott előnye van:

- Kicsi és áttekinthető: mivel csak a lényeget tartalmazza, kiváló kiindulási alapot jelent egy professzionális magyar helyesírás-ellenőrző program elkészítéséhez.
- A C++ nyelv megkönnyíti a keresztplatformos C++-alapú fejlesztésekhez való integrációját.
- Nagy mértékben hordozható, mivel eleve Unix/Linuxra és Windowsra fejlesztik egyidejűleg.
- Szintén a hordozhatósággal kapcsolatos, hogy a MySpell futásidőben állítja elő két egyszerű szöveges állományból – a ragozási szabálygyűjteményből (affixum állomány), és a szótárállományból – a működése során használt speciális adatszerkezeteket. Ezzel szemben az Ispell bináris adatállományokat használ, sőt az Ispell program fordítási idejű beállításaitól függ, hogy futásidőben képes-e feldolgozni ezt a bináris adatállományt, vagy sem.
- Egy Unix/Linux, és Windows operációs rendszerek alá készült professzionális irodai programcsomag, az OpenOffice.org része, amely szabadszoftverként minden felhasználó számára elérhető ingyenesen, jogtisztán és korlátozások nélkül.
- Az említett vonzó tulajdonságok miatt több közreműködő is folyamatosan fejleszti a MySpell-t. Így került bele a MySpell kódjába egy jelentős sebességoptimalizáció, egy a javaslattevés minőségét nagy mértékben növelő megoldás (javítási cseretáblázat), valamint a hiányzó szóösszetételképzési lehetőség és több kisebb hiba javítása.

- Az említett szóösszetétel-képzési kiegészítés javítja az Ispell egy bosszantó (bár kevésbé feltűnő) hibáját: az Ispell egy szó ragozott alakját akkor is elfogadja egy szóösszetételben, ha amúgy a tőszót nem (pl. kutyaokos hibás, de kutyaokosnak elfogadott szóösszetétel).

2.1.5. Mozilla Spell Checker

A MySpell forráskódján alapuló ellenőrző. David Einstein munkája. A Mozilla Spell Checker jelentős sebességoptimalizációja visszakerült a MySpell-be, és a tervek szerint a két ellenőrző egyesülni fog a közeljövőben.

2.1.6. Magyar MySpell

A MySpell forráskódján alapuló GPL licencű ellenőrző, amit 2002-ben kezdtem fejleszteni.

Először kisebb hibajelentések illetve foltok készültek a MySpell-hez, de később rákényszerültem az idő és a támogatni kívánt nyelvek sajátosságai ismeretének hiányában egy külön ág létrehozására.

A Magyar MySpell fejlesztés eredményeként került az eredeti MySpell ágba, és az így az OpenOffice.org-ba is a szóösszetétel-képzési lehetőség, valamint a javaslattevés minőségét jelentős mértékben emelő javítási cseretáblázat-kezelés.

A külön Magyar MySpell ág a következő fontosabb tulajdonságokkal rendelkezik:

- A magyar helyesírás szabályai 138. számú kötőjellel írási szabályához (3-6-os szabály) igazított szóösszetétel-ellenőrzés.
- Javaslattevés a szótárban nem szereplő szóösszetételek esetén is.
- A szókincs futásidejű bővítésének lehetősége.
- Az újonnan felvett szavak továbbragozása egy mintaként megadott tőszó alapján.
- Javított javaslattevés nagybetűt tartalmazó szavakra.
- A mássalhangzó-triplázások letiltása (sakkkör).
- Cső felület a beágyazáshoz.
- Ékezet nélküli szavak esetén helyes javaslattevés.
- Hibásan írt szavak helyes szóösszetételként való elfogadásának megakadályozása. Az egyes tőszavak szóösszetételben való megjelenésének (és ezzel teljes szóbokrok) letiltása helyett a javítási cseretáblázat (esetleg a teljes javaslattevő modul) felhasználásával megvizsgálható, hogy nem tipikus hibáról van-e szó ritka szóösszetétel helyett. (Ilyen hibás összetételek, amelyek korábban csak az egyik tőszó teljes letiltásával vagy sehogy nem voltak lekezelve: szervíz, vizitorma, birság, elitélt, vasárú, tejles, szívéjes, karvaj, stb.)

Számos egyéb tulajdonság megjelenése a közeljövőben várható:

- Kötőjeles és nagykötőjeles alakok ellenőrzése a kötőjellel kapcsolt toldalékok, ikerszók, a mozgószabály alapján egybeírt szóösszetételek, földrajzi nevek, számnevek és nagykötőjelet tartalmazó összetételek esetében.
- Szóismétlések kiszűrése (esetleg opcionálisan).
- Kötelező egybeírásra való figyelmeztetés (pl. legalábbis, véghezvisz).

2.1.7. OpenOffice.org

Nem kifejezett helyesírás-ellenőrző, de a Magyar MySpell fejlesztésének egyik kiemelt célpontja.

Az OpenOffice.org alapját képező StarOffice a német StarDivision cég kereskedelmi terméke volt a 90-es években. A Sun Microsystems a céget és termékét felvásárolta üzleti megfontolásokból. A cég 2000-ben GPL licenc mellett szabadszoftverré tette a StarOffice nagy (a StarDivision által fejlesztett) részét. A licencelt nyelvi eszközöktől, és adatbázistechnológiától megfosztott program az OpenOffice.org nevet kapta (az OpenOffice védett név volt, ezért a furcsa, a szabadszoftver projekt honlapjának címét hordozó elnevezés). A Sun jelenleg kereskedelmi támogatással vállalati ügyfelek számára az OpenOffice.org-on alapuló StarOffice-t is biztosítja.

Az OpenOffice.org jellemzői:

- Professzionális irodai programcsomag
- Kiváló Microsoft Office export, import szűrők
- Nagy méretű, sok képet tartalmazó dokumentumok gyors kezelése.
- GPL-s: ingyenes, módosítható, terjeszthető (illetve kettős licencelésű).
- XML alapú saját fájlformátum (Java jar formátumú állományokban).
- UNICODE támogatás, még ázsiai nyelvekhez is (mindkettő Sun fejlesztés)
- A GPL licenc alá nem helyezhető licencelt részek fokozatos kiváltása (*lingucomponent* kísérleti projekt: MySpell helyesírás-ellenőrző, ALTLinux elválasztó modul, szinonimaszótár, nyelvhelyességi ellenőrzés).
- Honosított: magyar helyesírás-ellenőrzés a Magyar MySpell helyesírás-ellenőrző, és -javító programmal, és a Magyar Ispell szótármodullal. Magyar elválasztás Mayer Gyula és Miklós Dezső `huhyph.tex` TEX elválasztási moduljával. Mintegy 20 000 programfelirat és -üzenet magyar fordítása.

A számos új funkció miatt is az OpenOffice.org sokáig fejlesztői stádiumban volt. Mivel azonban a program meglepően üzembiztos és jól használható, 2002-ben (marketing okokból is) kibocsátásra került az 1.0-s elnevezésű változat.

A fejlesztői változat jelenleg még az eredeti számozást használja (641D), de ez a számozás várhatóan a közeljövőben megszűnik.

A Magyar OpenOffice.org változatai a `http://office.fsf.hu` oldalon jelennek meg, a fejlesztői változatok a `http://office.fsf.hu/work` könyvtárban találhatók. A Magyar OpenOffice.org windowsos és linuxos változatai a OOo változatszám mellett egy csomagszámot is kapnak (egymástól függetlenül). Tehát pl. a legfrissebb linuxos Magyar OpenOffice.org teljes „neve”: Magyar OpenOffice.org 1.0, 43-as számú linuxos csomag.

2.2. Az Ispell felépítése és működése

Az Ispell adatállományai `.hash` kiterjesztéssel rendelkeznek, és Unix/Linux alatt a `/usr/lib/ispell/` vagy `/usr/local/lib/ispell/` könyvtárban kerülnek elhelyezésre.

Egy ilyen hash állomány (pl. a `magyar.hash`) két fő részből áll: tartalmazza a (ragozási) osztályokba sorolt ragozási szabályokat, és tartalmazza a tényleges hasítótáblát is. A hasítótábla tárolja a tőszavakat és minden egyes tőszónál még az azon alkalmazni kívánt ragozási osztály vagy osztályok kapcsolóit.

Egy Ispell hash-állomány a `buildhash` paranccsal állítható elő, amely két állományt vár bemenetként: egy ragozási szabályokat (és egyéb beállításokat) tartalmazó affixum állományt és egy szótárállományt, amely a tőszavakat és kapcsolóikat tartalmazza.

2.2.1. Ragozási szabályok

Az Ispell affixum állomány formátumát egy kézikönyvoldal (man 4 ispell) ismerteti részletesen. Itt most csak a ragozási szabályokat, és az affixum tömörítés módszerét vizsgáljuk meg közelebbről.

Egy ragozási szabály nem más, mint egy feltételes kifejezés, amely leírja, hogy milyen feltételek között milyen módon helyezhetünk toldalékot egy alapszóhoz.

Az Ispell ragozási szabályainak leírása emlékeztet a szabványos regexp reguláris kifejezésekre, és a következő fő jellemvonásokkal rendelkezik:

- A feltétel a tőszó utolsó (prefixum esetén első) legföljebb nyolc karakterére vonatkozik.
- A feltételben erre a legföljebb nyolc karakterre vonatkozóan szerepelhetnek karaktertartományok is.
- A toldalékolásnál a toldalék hozzáillesztése előtt lehetőség van a tőszó végének leválasztására. Itt viszont már nem használhatunk karaktertartományokat (pontosan ismerni kell a leválasztandó karaktert vagy karaktersorozatot).

Példák, amelyek bemutatják a ragozási szabályok szintaxisát. (A kettőskereszt után sor végéig tartó megjegyzés következik, amely már nem része a ragozási szabálynak.)

```
Z          >  Z         # ráz -> rázz
Z          >  ZÁL       # ráz -> rázzál
E D Z      >  ENEK      # edz -> edzenek, pedz -> pedzenek
[ÓUÚ]      >  VAL       # manó -> manóval, hamu -> hamuval, bú -> búval
[^AE]      >  NAK       # okos -> okosnak
.          >  KÉNT      # kutya -> kutyaként, okos -> okosként
A          >  -A,ÁNAK   # macska -> macskának
[^SD] Z I K >  -IK,Z    # barátkozik -> barátkozz
```

Az első két példa jelentése, hogy Z betűre végződő tőszó Z, és ZÁL toldalékkal rendelkező alakja is helyes (ez a kisbetűs alakokra is vonatkozik).

A harmadik példa szerint az EDZ betűsorozatra végződő szavak ENEK toldalékot is kaphatnak.

A negyedik példában karaktertartományt adunk meg az utolsó karakterre vonatkozóan: minden Ó, U, és Ú-ra végződő szó VAL toldalékot kaphat.

Az ötödik példa komplementer karaktertartományt ad meg: ha a szó nem A vagy E betűre végződik, megkaphatja a NAK toldalékot.

A hatodik példa a tetszőleges karakter jelét, a pontot mutatja be: bármely szó megkaphatja a KÉNT ragot.

A hetedik példa a tőszó végének levágását mutatja be: az A-ra végződő szavak ÁNAK toldalékot kaphatnak a szóvégi A levágásával.

Az utolsó egy kombinált példa: A nem SZIK, vagy DZIK végű, de ZIK-re végződő szavak Z toldalékot kaphatnak a szó végi IK levágásával. Így pl. a barátkozik→barátkozz helyes átalakítás, de a vakaródzik→vakaródzz, illetve mászik→mászz már nem.

2.2.2. Ragozási osztályok

Az előző példák meglehetősen keverve tartalmaztak különböző szófajú szavakra vonatkozó (és részben hibás) szabályokat. Az Ispell ragozási osztályok kialakításának lehetőségével biztosítja, hogy a tőszavakat a kategóriájuknak megfelelő ragokkal láthassuk el. A ragozási osztályok egy betűs (vagy egy karakteres) névvel rendelkeznek, amit a továbbiakban kapcsolóknak nevezünk.

Példa két ragozási osztály definiálására:

```
flag A:

[^L] L  >  LAL  # pedál -> pedállal
[^L] L  >  LÁ   # pedál -> pedállá
L L     >  AL   # szakáll -> szakállal
L L     >  Á    # szakáll -> szakállá

flag B:
```

```
L       >  NAK  # szaval -> szavalnak
L L     >  ANAK # javall -> javallanak
```

Az első osztály a mély hangrendű főnevek ragozásának egy részletét mutatja, az L betűre végződő szavak -val toldalékolását, míg a másik osztály a mély hangrendű alanyi igeragozás egy részletét.

2.2.3. Szótárállomány

Az Ispell szótárállományban (dict fájl) egyszerűen felsoroljuk a tőszavakat, a megfelelő kapcsolókkal ellátva:

```
macska/A/E/G
pedál/A/F
akar/B/C
szaval/B
legalábbis
```

Látható, hogy egy tőszó tetszőleges számú ragozási osztályba sorolható, de akár ragozási osztályt sem kötelező megadni.

2.2.4. Affixum tömörítés

Minden ragozási szabályhoz egy 256 byte-ból álló tömb kerül a hash állományba. A tömb szerepe, hogy gyorsan eldönthető legyen egy ragozott szóról, hogy az adott ragozási szabály alkalmazásával állt-e elő, vagy sem. A tömb az ASCII karaktertábla elemeivel indexelhető, és visszaad egy byte-ot, ami a már említett 8 karakterpozícióra vonatkozó előfordulási információkat tartalmazza: pl. ha a tömb['A'] elem értéke binárisan 11111010, akkor a tőszóra nem alkalmazható a ragozási szabály, ha a tőszó „A" betűre végződik, vagy a kettővel előtti pozíción is „A" betű található.

A MySpell forráskódja részletesen dokumentálva tartalmazza az affixum tömörítési algoritmust. Amiért itt említésre került, az az, hogy az affixum tömörítés elnevezés nem az affixum állomány kisebb méretre való tömörítésére utal, hanem arra, hogy a lehetséges gyors adatszerkezetek közül talán a legtömörebbet sikerült Geoff Kuenningnek megtalálnia, gyakorlatilag a ragozási szabályokban szereplő karaktertartományok hatékony tárolásával és kezelésével.

2.3. Magyar Ispell

A Magyar Ispell egy keretrendszer, amely magyar affixum és szótárállományt állít elő az Ispell számára. Az `i2myspell` program segítségével ezekből az állományokból hozható létre a megfelelő MySpell affixum és szótárállomány is.

2.3.1. Történet

A Magyar Ispell kialakítását 1998-ban kezdtem. 1999-re a tárgyas igeragozás, néhány képzős toldalékolás, sok határozószó és névutó kivételével a helyesírási-modul hibákkal ugyan, de alapjaiban elkészült. A félkész helyesírás-ellenőrzővel 1999-ben a Közművelődésügyi Minisztérium és a Matáv által közösen meghirdetett „A Magyarországi digitális kultúráért” pályázaton is részt vettem, de sikertelenül.

1999 őszén a helyesírás-ellenőrzővel való munkát befejeztem, az eredmények 2000 nyaráig, azonosítóm megszűntéig elérhetőek voltak a `http://sol.cc.u-szeged.hu/~nemethl/` címen.

2001 őszén Szántó Tamás, a KDE grafikus felület magyar honosítója felajánlotta a segítségét az ellenőrző fejlesztésének folytatásához. Tamás folyamatosan tesztelte az ellenőrzőt, számos (ragozási) típushibát megtalálva. A tesztelések eredménye volt a szókincs jelentős bővítése is.

Tamás munkáját Bíró Árpád folytatta, elsősorban további típus- és egészen különleges rejtett hibák feltárásával, a matematikai és egyéb nyelvi modulok összeállításával és a forrásállományok ellenőrzésével.

Godó Ferenc hatalmas szólistáival az ellenőrző alkalmassá vált a mindennapos munkára.

A Magyar Ispell fejlesztése 2002 januárjáig „titokban” folyt, ekkor jelent meg az első nyilvános változat.

A nyilvánosságra hozatalt követően csatlakozott hozzánk Trón Viktor, aki további hibák, illetve hiányosságok feltárása mellett internetes forrásokból összegyűjtötte, és a Magyar Ispell számára feldolgozta a mai és történelmi magyar helységneveket, illetve – aminek külön örülök – elsőként vállalkozott az alul dokumentált Magyar Ispell keretrendszer működésének megismerésére, és ennek eredményeképp a szótőgenerálás egyes részeinek javítására.

A Magyar Ispellből Koblinger Egmont – aki számos hasznos észrevétellel, és szólistával is hozzájárult a Magyar Ispellhez – készített folyamatosan (SuSE) RPM csomagokat, illetve Pásztor György Debian karbantartó DEB csomagokat.

Talán a Magyar Ispell projekt hatására a megalakulófélben lévő Magyar Szabadszoftver Alapítvány az OpenOffice.org honosítása mellett döntött, és szervezésükben egy hétvége alatt a programcsomag tízezres nagyságrendű feliratai, és üzenetei kerültek lefordításra. Jelenleg a magyar OpenOffice.org linuxos változatának Noll János, windowsos változatának Bencsáth Boldizsár a karbantartója (`http://office.fsf.hu/`).

2.3.2. Alapvető problémák

Az agglutináló nyelvekhez nem magától értetődő az Ispell szótármodulok elkészítése. A legjelentősebb megoldandó problémák a következőek voltak a Magyar Ispell kidolgozása során:

1. Az Ispellben nincs többszörös toldalékolási lehetőség. Míg a magyar nyelv gazdag a képzőkben, jelekben, és ragokban, addig az Ispell csak a tágabb értelemben vett ragok kezelésére képes. A lehetséges szó végi hasonulások, képzők, jelek és ragok kombinációi miatt

igen nagy számú ragozási szabály megadása szükséges, amit „kézzel” helyesen és belátható idő alatt megadni lehetetlen.

2. A szótárállományra ugyanez vonatkozik: közvetlen elkészítése nagy nehézségekbe ütközik, és az eredmény nehezen módosítható a későbbiekben.

3. Az egyes képzők olyan számú képző+jel+rag kombinációt állítanak elő, és ennek megfelelő számú ragozási szabályt igényelnek, amelyet már nem lehet elfogadható méretű állományban tárolni.

   A magyar Ispell hash állomány jelentős részét (több, mint 60%), a jelenlegi 10000 ragozási szabályt tároló affixum tömörítést használó adatszerkezet (tehát nem a tényleges szótári hasítótábla) teszi ki, mintegy 5-7 Mb-on. Ha az összes lehetséges képző kombinációt ragozási szabályokkal próbálnánk tárolni, nagy valószínűséggel nagyságrendi növekedésre lehetne számolni a fájlméretben és a helyesírás-ellenőrzés időigényében is.

4. Az Ispellben nincs lehetőség nem szótári tövek felvételére. A morfofonológiai alternációk egy része, mint a szó végi „a”, „e” változása (macska→macskát) jól kezelhető a ragozási szabályokkal, de a hangkivetős főnevek (eper→epret), a nyitótövek (kéz→kezek), a hangugratós igék (bérel→bérlem)[1] már nem, ráadásul ezek igen változatos formát mutatnak a magyar nyelvben.

5. Nemcsak a ragozási szabályok számát sem növelhetjük az égig, hanem a ragozási osztályok száma sem haladhatta meg pusztán gyakorlati megfontolásból a 26-ot (a legtöbb Linux terjesztés ugyanis pár éve ilyen korlátozásokkal fordított bináris Ispell-t tartalmazott).

Megoldás szerencsére mindegyik pontra akadt:

1. Az affixum állomány, és a szótárállomány is automatikusan van előállítva a Magyar Ispell keretrendszerben. Az affixum állományokban az `m4` makrófeldolgozó segítségével gyakorlatilag többszörös toldalékolással vannak leírva a ragozási szabályok. A `magyar.aff` névre hallgató Ispell affixum állomány a végleges ragozási szabályokkal fordítási időben jön létre.

2. A szótárállomány forrása jelenleg az alapmodulból és külön kezelhető szaknyelvi modulokból áll. A szavak osztályozása egyszerű szólistákkal („egymezős” adattáblákkal) történik, ami rendkívül leegyszerűsíti a szókincs bővítését a Magyar Ispell-ben. A szólistákból héj- és awk programok segítségével áll elő az Ispell buildhash programja számára szükséges magyar.dict szótárállomány.

3. A Magyar Ispell kombinált módszert használ a képzett szóalakok kezelésére: a ragozási szabályok mellett a főnevekhez képest viszonylag kis elemszámú szófajok – mint a melléknevek és az igék – esetében automatikusan állítja elő a képzett alakokat, és ezek külön kerülnek a szótárba más szófajú szavakként. Ezzel a módszerrel a szükséges ragozási szabályok számát jelentős mértékben csökkenteni lehetett, a szótárállomány növekedésének rovására. A szótárba külön felvett szóalakok számának növekedése mintegy kétszeres, de

[1] Lásd pl. [3, 832 skk.].

a hash állomány speciális felépítéséből fakadóan a végleges Ispell adatállomány mérete kevesebb, mint ötven százalékkal nő csak. (A ragozási szabályok számának nagyságrendi növelése ennél sokkal nagyobb problémát jelentene).

4. A nem szótári tövek kezelésére kényszerűségből szokatlan megoldást sikerült találni: a nem szótári tőhöz kapcsolódó alakok levezetése nem valódi tőből, hanem az egyik ragozott alakból történik. A módszer rendkívül eredményesnek bizonyult: alkalmas volt a magánhangzó-harmónia szerint elkülönülő kisebb, zárt főnévcsoportok ragozásának kezelésére is, valamint ezt és a különböző hangrendű alakok kezelését is egy ragozási osztályon belül el lehetett intézni (mivel a tőnek választott többes számú alak kötőhangzójából egyértelműen meghatározható a szó hangrendje)!

5. A ragozási osztályok számának csökkentése érdekében több látszólag nem optimális, de mint később kiderült, a fejlesztés szempontjából szerencsés választásra került sor: a ragozási osztályok egy része össze lett vonva, pl. az ikes és iktelen igék ragozása, a főnevek és melléknevek ragozása részben ilyen volt. A hangrend esetében sem a ragozási szabályok számával takarékoskodó négy osztályos megoldás mellett döntöttem (mély hangrend, közös magas hangrend, csak ajakkerekítéses, csak ajakréses magas hangrend), hanem a szabályokat jelentős részben duplázó, de eggyel kevesebb ragozási osztályt kívánó megoldás mellett (mély hangrend, és a két magas hangrend).

   Azóta szerencsére a 26 osztályos határt is sikerült átlépni, mindenféle probléma nélkül, mivel a Linux terjesztések is átálltak a több ragozási osztállyal fordított Ispell binárisokra.

2.3.3. Affixum keretrendszer

Az affixum állományok az áttekinthetőség miatt több részre vannak feldarabolva. A legfontosabb, amely a ragozási osztályok rövid összefoglalását is tartalmazza, az aff.fej névre hallgató állomány.

Az aff.alanyi és az aff.targyas állományok az alanyi illetve a tárgyas igeragozást tartalmazzák. Ezek az állományok hangrendi összevonást nem tartalmaznak, az m4 makrók, amelyek itt szerepelnek, csak a ragozási szabályokban szereplő mintákat általánosítják. Pl. az aff.alanyi állomány első (valóban is használt) makrójának definíciója a következő:

```
define(dupik,[BDGJKLMNRTZY] [LDGRZ] I K)
```

Egy példa a makró használatára, ugyanebből az állományból:

```
dupik     >   -IK,ASZ
```

Ez a sor az aff.alanyi állomány m4 makrófeldolgozón való átszűrésének eredményeképp a következőre alakul:

```
[BDGJKLMNRTZY] [LDGRZ] I K     >   -IK,ASZ
```

A makrók használata itt csökkenti a tévedéseket, egyszerűsíti a minták módosítását a későbbiekben.

Az aff.fonev állomány tartalmazza a főnevekre, és – az állomány megtévesztő neve ellenére – a melléknevekre vonatkozó ragozási szabályokat illetve ezek m4 forrását.

Az m4 makrózás használata ebben az állományban lényegesen megkönnyíti a számtalan rag előállítását. Az m4 makrók paraméteresek lehetnek, egymásba ágyazhatók, egyszerű feltételvizsgálatokat végezhetünk. Segítségével a mély és a két magas hangrend, valamint a tekintélyes számú képző+jel(ek)+rag kombináció rövid makrókkal legeneráltatható. A *pelda* állomány tartalma a bal oldalon, az m4-gyel feldolgozott állomány kimenete a jobb oldalon látható:

```
define(rag,                    Futtatás:
$1BA                           $ m4 pelda
$1BAN                          kocsiMBA
$1RA)                          kocsiMBAN
                               kocsiMRA
define(jel,                    kocsiDBA
rag($1M)                       kocsiDBAN
rag($1D)                       kocsiDRA
rag($1JÁ))                     kocsiJÁBA
                               kocsiJÁBAN
jel(kocsi)                     kocsiJÁRA
```

A *define(makró_név, makró_törzs)* szintaxis az m4 központi utasítása, a paraméterek $*szám* azonosítóval érhetők el. Részletes leírása: info m4.

További affixum állományok: aff.fonev.morfo, amely a hangkivetős és a nyitótövek ragozását tartalmazza (a többes számú alakból levezetve) két ragozási osztályt meghatározva. Az aff.ige_kiv a hangugratós igék ragozási osztályát definiálja.

2.3.4. Szótári keretrendszer

A szófajilag elkülönített szóállományok lehetővé teszik a szófajra jellemző ragozási osztályok kapcsolóinak és az alapszavaknak a viszonylag egyszerű párosítását. Héj- és awk program gondoskodnak erről bizonyos kivétellisták figyelembevételével.

Például egy szófajon belül három fő ragozási mintacsoport különíthető el a hangrend alapján. Az alapszavak hangrendje kibővített reguláris kifejezésekkel (awk programokkal) fordítási időben kerül felismerésre, pl. egy erre vonatkozó programrészlet:

```
# mély hangrendű igék
```

```
  /[uúoóaá][bcdfghjklmnprstvxyz]*ik$/ { print $0 "/X/A" }

  /[uúoóaá][bcdfghjklmnprstvxyz]*$/ && ! /ik$/ { print
$0 "/X/A" }
```

A példa csak az utolsó szótag hangrendjének vizsgálatát végzi el, illetve ikes igék esetén az -ik előtti szótagét, ez sok esetben nem ad jó megoldást. A Magyar Ispell keretrendszerben kivétellisták segítségével tehető pontossá az automatikus hangrend-besorolás. Ilyen kivétellisták az `_alap` könyvtárban található `fonev_mely`, `ige_mely`, és `melleknev_mely` állományok, amelyeket a feldolgozó awk programok egy-egy asszociatív tárba töltenek be (más kivétellistákhoz hasonlóan), ami a feldolgozás sebességét jelentősen gyorsítja.

2.4. Új szótármodulok létrehozása

A Magyar Ispell forrásának telepítése után (ezt a következő fejezet ismerteti) lépjünk be a `magyarispell-0.86` könyvtárba.

A szótármodulok külön könyvtárakban helyezkednek el. Az alapkönyvtárak neve aláhúzásjellel (_) kezdődik. A legfontosabb szótármodul az `_alap`.

Egy új szótármodul létrehozása nagyon egyszerű: hozzunk létre egy új alkönyvtárat a `magyarispell-0.86` könyvtáron belül. A könyvtár neve aláhúzásjellel kezdődjék.

Ha szeretnénk, hogy a szótár fordítása során az új szótári modul is bekerüljön a végeredménybe, hozzunk létre egy szimbolikus kötést (symlink) a könyvtárra:

```
ln -s _példa Példa
```

Látható, hogy a többi alkönyvtárhoz hasonlóan a szimbolikus kötés neve aláhúzás nélküli, és nagybetűvel kezdődik. Ezek a tulajdonságok határozzák meg, hogy mely könyvtárak tartalma számít fordításra kijelölt modulnak.

Egy modult egyszerűen a szimbolikus kötés törlésével tudunk kivenni ebből a csoportból:

```
rm Példa
```

2.4.1. Szófaji bontás

A szótármodul szavakat fog tartalmazni, amelyet szófaji alapon el kell tudnunk különíteni. A köznevek a `fonev` állományban kerülnek felsorolasra (egy szó egy sorba kerül). A melléknevek a `melleknev` állományba, illetve az alanyi és tárgyas igék az `ige_alanyi` és az `ige_targy` állományokba.

A `tulajdonnev` állomány formátuma elsőre kicsit zavaró lehet: ha egy tulajdonnév nem kap toldalékot a magyar nyelvben (mozaikszó, vagy a szó vége idegen), akkor a sorban egy tabulátor után (tehát a második mezőbe) írjuk a szavakat.

```
Bastille
```

Ha ragozzuk, akkor rögtön sor elejére

```
Einstein
```

Ha a ragozott alakok között melléknévképzős (kisbetűs) is akad, akkor az ugyanebbe a sorba, egy tabulátorral elválasztva írjuk:

```
Einstein einsteini
```

Ha sehová nem illik a szó, felvehetjük a `ragozatlan` állományba. Ha azért valamilyen mértékben ragozott, akkor puskázzunk. A legenerált `magyar.dict` állományt szűrjük meg arra vonatkozóan, hogy egy hasonló ragozású, hangrendű, stb. szó milyen ragozási csoportjelzőkkel került a szótárba:

```
grep  ^milyen/ magyar.dict
milyen/B/V/L/R
```

A szavunkhoz ugyanezt rendelhetjük hozzá a `ragozatlan` állományban:

```
ilyen/B/V/L/R
```

2.4.2. Kivételek a szófaji kategóriákon belül

A legtipikusabb kivételek a hangrendi kivételek. Ha egy szó utolsó szótagja magas hangrendű, de a szó mégis mély hangrendű, akkor a szót vegyük fel a szófajhoz tartozó `_mely` végződésű névvel rendelkező állományba. Az ilyen szavak utolsó szótagjában többnyire „é”, vagy „i” magánhangzó szerepel, és gyakran vegyes hangrendűek.

Konvenció, hogy `tulajdonnev_mely` állomány nem létezik, az ilyen tulajdonneveket a `fonev_mely` állományba helyezzük el.

2.4.3. További információk

További információkat a Gyakran ismételt kérdések között találunk (`GYIK` állomány a Magyar Ispell forrásában), valamint a Magyar Ispell levelezőlistán (`http://www.yahoogroups.com/group/magyarispell`).

A levelezőlistára a `magyarispell-subscribe@yahoogroups.com` e-mail címre írt levéllel jelentkezhetünk. A listára szánt leveleinket a `magyarispell@yahoogroups.com` címre címezzük.

2.4.4. Segédprogramok

A `bin/` alkönyvtárban található `break` program segít szavakat kigyűjteni szöveges, vagy HTML állományokból.

2.5. Magyar MySpell

A Magyar MySpell legfrissebb változata is a Magyar Ispell projekt honlapjáról, a `http://www.szofi.hu/gnu/magyarispell` oldalról tölthető le.

A forrásból a `make` paranccsal fordíthatjuk le az `example` nevű példaprogramot, és a leendő – Ispell-t felváltani képes – programot, a `Mispell`-t.

2.6. Javítási cseretáblázat

A következő táblázat összefoglalja a Magyar Ispell 0,86-os változatában megtalálható javítási cseretáblázatot.

A táblázat harmadik oszlopa részben példákat mutat a cseretáblázat segítségével tett javaslatokra, valamint a cseretáblázat segítségével beazonosított hibás szóösszetételekre. Ez utóbbi tulajdonság a Magyar MySpell 0,4-es változatában jelent meg.

Mit	Mire	Példák cserére, illetve javított hibákra
í	i	szer+víz, sí+ma
i	í	vizit+orma, elit+élt
ó	o	mikró+számítógép
o	ó	
ú	u	lúd+as, vas+árú
u	ú	
ű	ü	
ü	ű	
j	ly	juk→lyuk, est+éj, kar+vaj, kord+éj, rost+éj, szeg+éj, szem+éj, szem+éjes, szem+éji, szent+éj, szesz+éj, szesz+éjes, tar+táj, tök+éj, ünnep+éj, ünnep+éjes,
ly	j	csevely→csevej, muszály→muszáj, boly+torján, kard+bolyt, tojássár+gálya, váll+bolyt, boly+ár, boly+tár, súly+tó
jj	lly	gajj→gally
jj	llj	ájj→állj
lly	jj	csevellyel→csevejjel
ggy	gyj	haggyon→hagyjon, naggya→nagyja
gyj	ggy	higyjen→higgyen

ggy	dj	maraggy→maradj
gy	dj	horgya→hordja, kargyuk→kardjuk
nny	nyj	annya→anyja
nny	nj	fonnyátok→fonjátok
tty	tyj	attya→atyja
tty	tj	láttyák→látják, bottya→botja
cc	tsz	szerecc→szeretsz
cc	dsz	maracc→maradsz
cc	gysz	eccer→egyszer
cs	ts	kölcség→költség
cs	ds	boloncság→bolondság
ccs	ts	baráccság→barátság
ccs	ds	vaccság→vadság
ccs	gys	naccság→nagyság, őnaccsága→őnagysága
g	kd	csug→csukd, lög→lökd
öss	ős	elösször→először, erössen→erősen
öll	ől	szöllő→szőlő
ütt	űt	hüttő→hűtő, füttő→fűtő
ijj	íj	dijja→díja, ijja→ídja
x	ksz	boxer→bokszer, vax→vaksz
ksz	x	fakszol→faxol, Beatriksz→Beatrix
xx	xsz	kódexxel→kódexszel
xx	kssz	vaxxal→vaksszal
xsz	kssz	vaxszal→vaksszal
x	xsz	kódexel→kódexszel
x	kssz	vaxal→vaksszal
sz	c	licensz→licenc
szt	cet	licenszt→licencet
ssz	cc	licensszel→licenccel
jl	lj	tejles→teljes, éjlen→éljen

3. fejezet

Telepítés és használat

A Magyar Ispell projekt eredményei számos szoftverrel működnek együtt, ezért a telepítés ismertetése is ezekhez a szoftverekhez kapcsolódik.

3.1. Az Ispell telepítése

3.1.1. Linux

Az Ispell része a legnagyobb Linux terjesztéseknek, és egy átlagos Linux telepítésnél az Ispell telepítésére is sor kerül automatikusan.

Linux alatt az Ispell szótárállományai a /usr/lib/ispell állományban kerülnek elhelyezésre. A Magyar Ispell telepítése során a magyar hash (`magyar.hash`) állomány is ide kerül.

Telepítés forráskódból

Az ispell-honlapról (`http://fmg-www.cs.ucla.edu/fmg-members/geoff/tars/`) letölthető forráscsomag a benne lévő `README` fájlban leírtak szerint (make all ; su ; make install) lefordítható. A local.h fájl linux esetében javasolható módosításai a

```
#define NO8BIT    /* Remove this if you use ISO character sets */
```

sor törlése és a

```
#define USG       /* Define on System V or if term.c won't compile */
#define LIBDIR    "/usr/lib/ispell"
```

sorok eredetijeinek átírása a közöltekké.

Némely rendszerek esetében szükség lehet a `/usr/lib/` könyvtárban a `# ln -s libtermcap.so.2 libtermcap.so` kötés létrehozására.

Ezután már csak make és – rootként – make install, mely parancs a `/usr/lib/ispell/`-be teszi az angol adatfájlokat és /usr/local/bin/-be a programokat.

3.1.2. Windows

A Windowsra történő telepítéshez Geoff Kuenning Ispell honlapján (`http://fmg-www.cs.ucla.edu/fmg-members/geoff/ispell-winnt.html`) több lehetőséget is találunk.

Windows alatt két fő módja van az Ispell telepítésének és használatának. Az elsőben a teljes Cygwin környezetet telepítjük a Windows NT, 2000, illetve XP operációs rendszerünkre, ami gyakorlatilag teljes Unix környezetté egészíti ki a Windowst, és utána telepítünk egy Cygwin környezet alá fordított Ispell változatot.

A másik módszer azok számára, akik a szokásos Windows környezetet előnyben részesítik, egyszerűbb: ez a bináris csomag formájában letölthető WIN32-re fordított Ispell változat Cygwin nélküli használata.

Mindkét módszer esetében az `ftp://ftp.franken.de/pub/win32/develop/gnuwin32/cygwin/%porters/Humblet_Pierre_A/V1.1/` címen található verziót javasoljuk. A telepítést az `ispell-3.2.06-cygwin-1.3-bin.README` nevű fájl ismerteti. Az egyszerűbb esetben a csomagból kizárólag a `usr/local/bin/ispell.exe` fájlra van szükség, melyet a PATH-ban található valamelyik könyvtárba kell elhelyezni. A szótárfájlok (lásd 3.2.3) a `DICTDIR` környezeti változó által meghatározható helyre teendők. (A .tar.gz végződésű fájl kicsomagolásához a Cygwin környezeten kívül pl. a djtarnt.exe program használható.)

Ha tehát a `C:\AUTOEXEC.BAT` fájlban a

```
SET PATH=C:\WINDOWS;C:\WINDOWS\COMMAND;C:\BIN
SET DICTDIR=C:\ISPELL
```

sorok találhatók, akkor az ispell.exe fájlt a `c:\bin`, a magyar.hash és a magyar.aff fájlt a `c:\ispell` alkönyvtárba kell helyezni.

Szintén a PATH-ban kell elhelyezni a cygwin1.dll nevű fájlt.

Egy másik bináris csomag formájában letölthető WIN32-re fordított Ispell változat található a `ftp://ftp.tue.nl/pub/tex/GB85/ispell-dutch96/ispellw32.zip` címen.

A letöltés után csomagoljuk ki a ZIP állományt a `C:\` meghajtó gyökerében, és adjuk a keresési útvonalunkhoz a `C:\Ispell\bin` könyvtárat. Ehhez nyissuk meg a Start menü→Beállítások→Rendszer párbeszédablakot, és a Speciális lapon a középső részben (Környezet) keressük meg a keresési útvonalakat rögzítő környezeti változót (PATH), és írjuk a végére pontosvesszővel elválasztva a `C:\Ispell\bin` útvonalat.

Sajnos ez a dán Ispell modulhoz készült csomag még kis számú ragozási osztállyal lett lefordítva, ezért ha a legfrissebb Magyar Ispell szótármodult szeretnénk használni vele, újra kell fordítanunk az Ispellt. Ehhez az EMX+gcc csomagokra van szükség.

Egy harmadik win32-változat a CTAN:fptex (`ftp://ftp.dante.de/tex-archive/systems/win32/`) disztribúcióban található.

3.2. A Magyar Ispell telepítése és fordítása

3.2.1. Elérhetőség

A `http://www.szofi.hu/gnu/magyarispell/` oldalról töltsük le a legfrissebb változatát a programnak (`magyarispell-0.86.tar.gz`).

A már lefordított kész modulok részei a Linux terjesztéseknek, illetve a Magyar OpenOffice.org már tartalmazza is a Magyar Ispell/MySpell egy korábbi változatát.

3.2.2. Az Ispell változat megfelelősége

Ahogy már említésre került, az Ispell számos fordítás során beállított paramétertől függően végzi tevékenységét. Különösen régebbi Linux terjesztésekben, vagy Windows alatt meg kell győződnünk arról, hogy az Ispell képes helyesen lefordítani a Magyar Ispell forrásállományait.

Adjuk ki Linux/Unix alatt a következő parancsot:

```
ispell -vv | grep BIT
```

Ha a bináris Ispell megfelel a céljainknak, akkor a következő két sort kell látnunk:

```
MASKBITS = 64
!NO8BIT (8BIT)
```

Az első sor jelzi, hogy elegendő számú kapcsolót kezel az Ispell (a régebbi Ispell változatoknál a 32 volt az alapértelmezett, ami durván az angol ábécé nagybetűinek használatát tette lehetővé.)

A második sor jelzi, hogy a szótárban található szövegek kiterjesztett ASCII kódkészlettel (karakterenként 8 biten, és nem 7-en) vannak leírva.

Windows alatt a következő parancsot adjuk ki a parancssorban:

```
ispell -vv | more
```

A listában keressük meg a fenti két sort. Ha a MASKBITS értéke 32, vagy a NO8BIT értéke igaz, akkor a Magyar Ispell használatához szükség van az Ispell újrafordítására (vagy használjunk Magyar OpenOffice.org-ot).

3.2.3. Telepítés

A keretrendszer Linux/Unix, vagy Windows NT/Cygwin környezetben futtatható le. A letöltött állomány csomagolása, lefordítása, és telepítése:

Fontos! Az utolsó parancsot csak rendszergazdaként adhatjuk ki, mivel az a tényleges helyesírási-modult – a fordítás végeredményét – a `magyar.hash` állományt a `/usr/lib/ispell` könyvtárba másolja át.

```
tar xzvf magyarispell-0.86.tar.gz
cd magyarispell-0.86
make all
make install
```

A telepítéshez és a fordításhoz kb. 16 Mb szabad hely szükséges a háttértárolón. A fordításnál előálló `magyar.hash` állomány mérete Ispell változattól függően 11–13 Mb között változhat.

MSWindows használata esetében a hash- és az aff-fájl a 3.1.2 pontban leírt helyre teendő.

3.3. Magyar OpenOffice.org szótárállományok frissítése

3.3.1. Linux környezetben

A `make all` parancs hatására az OpenOffice.org (illetve az abban található MySpell helyesírás-ellenőrző) számára szükséges két állomány is létrejön a `magyarispell-0.86` könyvtárban.

Az állományok neve `hu_HU.aff`, és `hu_HU.dic`.

Mozgassuk át a két állományt az OpenOffice.org telepítési útvonalára, a `user/wordbook` alkönyvtárba:

```
mv hu_HU.* ~/OpenOffice.org1.0/user/wordbook/
```

Ezután újraindítva a Magyar OpenOffice.org-ot, az alkalmazás a legfrissebb szótármodult tölti be, és használja.

3.3.2. Windows alatt

Az OpenOffice.org számára szükséges két állomány külön is letölthető a `http://www.szofi.hu/gnu/magyarispell/` oldalról.

A két állományt tartalmazó tömörített ZIP állomány neve: `hu_HU0.86.zip`.

Letöltés után csomagoljuk ki az állomány tartalmát a Magyar OpenOffice.org telepítési útvonalára, a `user/wordbook` almappába.

Ezután újraindítva a Magyar OpenOffice.org-ot, az alkalmazás a legfrissebb Magyar Ispell szótármodult tölti be, és használja.

3.4. Az Ispell használata

3.4.1. Szöveges állományok

A következő paranccsal kérhetjük szövegfájlok ellenőrzését:

```
ispell -d magyar fájl(ok)
```

(Linux esetében egy kötés a hungarian név használatát is lehetővé teszi.)

Ebben az esetben megjelenik egy kényelmes és gyors menü, ami sorban felmutatja a hibásnak vélt szavakat.

Ha van javítási javaslat, akkor a javaslat előtt látható számjegyek leütésével fogadjuk el a javaslatot.

A legfontosabb billentyűparancsok a következőek:

- szóköz: Átugrik a szón, azt változatlanul hagyva.
- számok: Választás a javasolt cserék közül.
- *r*: Csere az általunk megadott szóra.
- *a*: Elfogadjuk a szót az ellenőrzés befejeződéséig.
- *i*: Felvesszük a saját szótárunkba. (Ez a `~/.ispell_magyar`, vagy ha létezik az aktuális könyvtárban az `.ispell_magyar` állomány, akkor az. Később – ha elhagytuk az Ispell programot – ez akár kézzel is módosítható.)
- *u*: Felvesszük a saját szótárunkba, de kisbetűsre alakítjuk előtte.
- *x*: A következő fájlra ugrunk. A félig ellenőrzött állománynak a fennmaradó része nem lesz kijavítva.
- *l*: A rendszerszótárban végezhetünk kereséseket. A `*` jel segítségével mintázatot is megadhatunk, hasonlóan a fájlrendszerben megadott útvonalakhoz: pl. az `L*` mintázat minden L betűvel kezdődő szót kilistáz a `/usr/dict/words` állományból. A rendszerszótár angol nyelvű.
- *q*: Kilépés mentés nélkül, ha a „q” után „y”-t nyomunk megerősítő kérdésre. A vizsgált állomány javításai elvesznek.

3.4.2. TEX állományok

A következő paranccsal kérhetjük TEX, illetve LATEX állományok ellenőrzését:

```
ispell -t -d magyar fájl(ok)
```

Amennyiben az állomány(ok) kiterjesztése .tex, a -t kapcsolót nem szükséges megadni.

A -t kapcsoló hatására az Ispell nem törődik a fordított perjellel kezdődő, ékezetes betűket nem tartalmazó szavakkal (vagyis a TeX parancsokkal). Bizonyos parancsokat felismer, és az utána következő egy, vagy két argumentummal sem foglalkozik. A matematikai környezetben lévő szavakat nem ellenőrzi az Ispell.

3.4.3. HTML állományok

A következő paranccsal kérhetjük HTML állományok ellenőrzését:

```
spellhtml -d magyar fájl(ok)
```

A spellhtml a Magyar Ispell feltelepítése során a /usr/bin-ben elhelyezett héjprogram (plusz spellhtmlic névre hallgató kiegészítő szűrő), ami az ellenőrzés előtt megfelelő (gyakorlatilag TEX-nek látszó) formátumra hozza az állományokat. Az ellenőrzés befejeztével visszaalakítja HTML állománnyá.

3.5. A Magyar OpenOffice.org telepítése

Az OpenOffice.org magyar változata a `http://office.fsf.hu` weboldalról tölthető le Windows, és Linux platformra is, illetve számítástechnikai lapok mellékletén (pl. a Chip Magazin 2002. júliusi száma) is megtalálható.

A program saját grafikus és magyar nyelvű telepítővel rendelkezik. Kövessük a honlapon, illetve a lapban található útmutatásokat!

3.6. Az Emacs telepítése

Linux esetében a disztribúciók tartalmazzák az Emacsot.

Mswindows-ra több helyen is találhatunk bináris csomagot, pl. az `ftp://ftp.gnu.org/gnu/windows/emacs/latest/` címen. Általános esetben az `emacs-21.2-bin-i386.tar.gz` csomagra van szükség, melyet pl. az `ftp://ftp.gnu.org/gnu/windows/emacs/utilities/i386/` címen található programokkal lehet kibontani.

Egy másik lelőhely a CTAN archívum, pl.: `ftp://ftp.dante.de/tex-archive/systems/win32/fptex/0.5/support/emacs-21.1.1-win32.zip`.

3.7. Emacs integráció

Az Emacs program az Ispell (illetve a MySpell) cső felületét kihasználva, a háttérben indított Ispell programmal kommunikálva képes a szövegek helyesírás-ellenőrzésére.

Az Emacs-ban való használathoz szükségünk van az `ispell.el` Emacs-lisp programra is. Ez még hivatalosan nem tartalmazza a magyar szótárra vonatkozó bejegyzést, ezért a módosított `ispell.el` állományt a `http://www.szofi.hu/gnu/magyarispell` Magyar Ispell honlapról tölthetjük le.

Az `ispell.el` állományt másoljuk abba a könyvtárba, ahol a lefordított `ispell.elc` található (pl. `emacs-21.1/lisp/textmodes/` könyvtár, régebben pl. `/usr/share/emacs/20.7/lisp/`).

Ezután vagy ki kell törölni a régi ispell.elc fájlt, vagy fordítással újat kell létrehozni (és ezáltal felülíratik a régi).

A fordítás – általában root jogosultsággal – legalább háromféleképp elvégezhető:

a) parancssorból a megfelelő könyvtárban adjuk ki az `emacs -batch -f batch-byte-compile ispell.el` utasítást

b) emacs-ban adjuk ki az `M-x byte-compile-file` parancsot, és adjuk meg az ispell.el fájl elérését

c) nyissuk meg az Emacs-ban a megfelelő helyre bemásolt `ispell.el` állományt és használjuk a menüsor Emacs-Lisp→Byte-Compile This File pontját

A figyelmeztetések ellenére a fordítás sikeres lesz, és a Magyar Ispell elérhetővé válik az Emacs programon belül.

3.8. Helyesírás-ellenőrzés az Emacs-on belül

Adjuk ki az `M-x ispell-change-dictionary` parancsot, és válasszuk ki a magyar szó beírásával a magyar Ispell szótármodult.

A teljes szöveg ellenőrzésére szolgál az `M-x ispell` parancs. Egy szót az `M-x ispell-word` paranccsal ellenőriztethetünk. Ehhez létezik gyorsbillentyű is: `M-$`.

Ha magyar az alapbeállításunk, természetesen az emacs indító konfigurációjába (vagyis a `~/.emacs` fájlba) is beleírhatjuk:

```
(ispell-change-dictionary "magyar")
(ispell)
(ispell-kill-ispell)
```

A folyamatos ellenőrzés alatt használható billentyűparancsok megegyeznek az Ispell billentyűparancsaival.

Minthogy a Spell-menü tartalmának alapbeállítása az emacs programba bele van fordítva, indításkor az nem tartalmazza a magyar lehetőséget, viszont az ispell legalább egyszeri indítása után már igen, tehát ekkor már a menü is használható a parancssori utasítások helyett: a 20. verzió esetében Edit→Spell, a 21. verzió esetében pedig Tools→Spell Checking.

3.9. Helyesírás-ellenőrzés beírás közben

Az Emacs *flyspell* üzemmódja lehetőséget nyújt a beírás közbeni helyesírás-ellenőrzésre. Az üzemmódba az `M-x flyspell-mode` paranccsal léphetünk be.

Ezután a hibás szavak grafikus felületen piros színűvé és aláhúzottá válnak. Ha a 3. egérgombbal (vagy a két szélsővel egyszerre) kattintunk a szón, felnyílik egy menü, amivel a javasolt cseréket fogadhatjuk el, illetve továbbléphetünk, felvehetjük a hibás szót az elfogadott szavak közé az ellenőrzés idejére, illetve „végérvényesen”. Ez utóbbi esetben a felvett szavak az Ispell program `.ispell_magyar` állományába kerülnek.

3.10. A Magyar Myspell telepítése

A http://www.szofi.hu/gnu/magyarispell/myspellhu-0.5.tar.gz címen elérhető a Magyar MySpell 0.5-ös változata.

A végrehajtható program neve mispell.

A könyvtárfájlok helye a mispell.cxx fájlban a LIBDIR makró segítségével adható meg; a jelen beállítás /usr/lib/myspell/.

A /usr/lib/myspell könyvtárban kötést kell létrehozni magyar.aff, magyar.dic néven a `hu_HU.aff` és `hu_HU.dic` állományokra. Amennyiben alapértelmezésben Az angolhoz default.aff, default.dic néven kötendők a megfelelő fájlok.

Az Emacs-szel/csal való használat legegyszerűbben úgy valósítható meg, hogy a mispell programmal fölülírjuk az ispell programot, és minden mást az ispellnél leírtak szerint teszünk.

A személyes szótárat a `$HOME/.ispell_`*nyelv*, vagyis pl. a `~/.ispell_magyar` fájlba írja a program.

A mispell a LyX-szel is együttműködik.

Bizonyos adatokat a /tmp/mispell.log fájlba halmoz.

3.11. A dokumentáció

A dokumentáció forrása LaTeX nyelven készült. A mesterfájlban a `\usepackage{times}` utasítás abból a célból szerepel, hogy a `pdftex` program segítségével jó minőségű és kis méretű pdf-változat lehessen generálható.

A dokumentáció html-változata a

```
latex2html -split 4 magyarispell
```

parancs segítségével szinte hiba nélkül generálható.

3.12. Felhasználási engedély

A Magyar Ispell jelen változatát Németh László készítette a Magyar Ispell projekt résztvevőinek segítségével, és a GNU GPL (General Public License) kettes, vagy későbbi változata alapján szabadon felhasználható.

A Magyar MySpell jelen változatát Németh László készítette, és a GNU GPL (General Public License) kettes, vagy későbbi változata alapján szabadon felhasználható.

A programok és a dokumentáció elkészítéséhez a TypoTEX Kft. 375 000 forinttal járult hozzá, a Széchenyi Terv keretében az Informatikai Kormánybiztosság SZ-IS-10/3 pályázati számon elnyert támogatásából.

4. fejezet

A Magyar Ispell tesztelése

A Magyar Ispell működésének tesztelésére egy független szókincstár, a Huhyph 4.0 elválasztási szóadattár szolgált.

A tesztrendszer két tesztelést végez el: az első a Huhyph szókincstár felismerését teszteli, a második a HuHyph-ból kinyert véletlen mintából különféle algoritmusokkal előállított hibás alakok helyesként való elfogadását teszteli.

4.1. A szókincs tesztelése

Csomagoljuk ki a `magyarispell-teszt.zip` állományt, és a könyvtárba lépve adjuk ki a `make teszt1` parancsot.

A szóadat szavai kötőjellel jelölve tartalmazzák az elválasztási pontokat, valamint fordított perjellel bevezetve, és kapcsos zárójelpárral lezárva a kettőzött többjegyű mássalhangzókat. Ezeket a jeleket az 1. teszt során töröljük, és a Magyar MySpell függvénykönyvtárat használó `badwords` programmal kiszűrjük az állományból a hibás szavakat.

Jelenleg a Magyar Ispell szótármodullal a `huwords.hyph` állomány 98,5%-a helyesnek bizonyul.

A maradék részben (ami mintegy 900 szót jelent a huwords.hyph kb. 65 000 szavából) még feldolgozás alatt álló, részben pedig idegen szavakat tartalmaz. Ha ezek elfogadását szeretnénk, az első teszt futtatásánál kapott `GEN.nemismert` állományt vegyük fel a Magyar Ispellben új szótári modulként, `ragozatlan`-ra átkeresztelve az állományt.

4.2. Tévesztések ellenőrzése

A tesztelés megkezdése előtt a `teszt.sh` állomány `JAVA_PATH` változójának adjuk meg a `java` virtuális gép helyét, mivel a téves alakokat előállító program (`MITest`) Java nyelvű.

Adjuk ki a `make teszt2` parancsot a tévesztések elfogadásának tesztelésére.

A tesztprogram a huwords.hyph szavaiból egy véletlen, 1000 szóból álló mintát vesz. Ebből a mintából a `MITest` program téves alakokat képez az alapvető Ispell javítási algoritmusokkal (betűkihagyás, betűbetoldás, betűcsere, szomszédos betűk felcserélése).

A teszt során három állomány keletkezik: A `GEN.osszes` állomány tartalmazza az összes előállított téves alakot. Ez mintegy 750-szerese az eredetileg 1000 szavas véletlen mintának.

A `GEN.osszes` állományból a `badwords` program kiszűri a helytelennek felismert szavakat, ez kerül a `GEN.hibas` állományba.

Végül ezeket a szavakat kivonjuk a `GEN.osszes` állományból, így kapjuk meg a `GEN.helyes` állományt.

A `GEN.helyes` tartalmazza tehát a `huwords.hyph`-ból nyert véletlen minta „elrontásából” keletkezett, mégis helyesnek felismert szóalakokat.

4.3. A tesztek értékelése

4.3.1. Szókincs

A `huwords.hyph` első feldolgozása során az eredeti állomány mintegy 5%-a, 3700 szó nem került felismerésre.

A listában az idegen, régies, ill. ritka szavak mellett a hiányzó tövek ragozott alakjai is előfordultak, így mintegy 2000 db. hiányzó tőről beszélhetünk, amely szám – tekintve a szavak többnyire nem hétköznapi jellegét – jónak mondható!

A hiányzó szavak listája nagyrészt feldolgozásra került Bíró Árpád munkája révén, így sikerült az 5%-ot jelentősen lecsökkenteni.

4.3.2. Tévesztések

A `GEN.helyes` szólista a többszörös tesztelések során átlagosan 16-szor bizonyult nagyobbnak, mint a véletlen minta.

Íme az „álomszerű” szó téves alakjaiból előállt helyes alakok:

```
álomászerű
álomészerű
álomízerű
álomőszerű
álomőzerű
álomsíerű
álomsóerű
álomszedrű
álomszegű
álomszelű
```

```
álomszemű
álomszenű
álomszer
álomszerb
álomszerbű
álomszere
álomszeré
álomszerfű
álomszeri
álomszermű
álomszert
álomszertű
alomszerű
álomszerű
álomszérű
álomszerűr
álomszerűt
álomszerv
álomszervű
álomszexű
álomszírű
álomszóerű
álomszőrű
álomszúerű
álomszűrű
```

Látható, hogy a szavak nagy része szokatlan szóösszetételként került elfogadásra.

Az ilyen szóösszetételek a Magyar MySpell 0.4-es változatában részben már nem kerülnek elfogadásra. Szerencsére erre a legtipikusabb hibák esetében (pl. i/í, o/ó, u/ú, ü/ű, j/l, j/ly, stb.) kerül sor, így a Magyar Ispell/MySpell már (ellentétben pl. a Helyes-e?-vel) nem fogadja el az ilyen típusú, tipikus és súlyos tévesztéseket (pl. szervíz, szeméj, tejles, stb.).

A teljes listából kiderül a Magyar Ispell/MySpell pár ismert hibája is: a nem összetett számnevek előfordulhatnak még bárhol a szóösszetételekben, illetve a hangkivetős főnevek, valamint a nyitótövek többesszáma szintén. Ezek javítása a közeljövőben várható.

Irodalomjegyzék

[1] JAKAB László: Tanulmányok az igeragozás köréből. KLTE Nytud. Int. 73. Debrecen, 1999.

[2] KIEFER Ferenc, szerk.: Strukturális magyar nyelvtan 2. Fonológia. Akadémiai Kiadó, Budapest, 1994.

[3] KIEFER Ferenc, szerk.: Strukturális magyar nyelvtan 3. Morfológia. Akadémiai Kiadó, Budapest, 2000.

[4] MAGYAR TUDOMÁNYOS AKADÉMIA, szerk.: A magyar helyesírás szabályai. 10. kiadás. Budapest, 1954, 1973[13].

[5] MAGYAR TUDOMÁNYOS AKADÉMIA, szerk.: A magyar helyesírás szabályai. 11. kiadás. Budapest, 1984, 1994[11].

[6] NÉMETH Anikó: A magyar nyelvtan. Merényi Könyvkiadó, Budapest, 1997.

[7] PAPP Ferenc, szerk.: A magyar nyelv szóvégmutató szótára. Akadémiai Kiadó, Budapest, 1969.

[8] TÖBBEK. ispell manpages (man ispell, man 4 ispell).