Magyar nyelvű tipográfia LATEX-hel

Szabó Péter <pts@fazekas.hu>

BME Számítástudományi és Információelméleti Tanszék

Második magyar T_EX-konferencia 2005-12-03 11:30 Budapest, ELTE déli tömb



Honosítás (L10N)

Áttekintés

Honosítás (L10N)

Magyar szöveg készítése L^AT_EX-hel Ki mit tesz Magyar L^AT_EX Ábra Bábelben

Gyakori hibák

Összefoglalás

A LATEX mint szoftver honosítása ma:

- *felhasználói felület*: nem történt meg, a rendszerparancsok, a T_EX-parancsok és a hibaüzenetek is angol nyelvűek. Egyes szövegszerkesztők felhasználói felülete már magyar.
- súgó: a L^AT_EX-nek nincs súgója.
- *telepítő*: sokféle T_EX-disztribúció létezik, az ismertebbeknek nincs magyar telepítője (kivétel: a teTeX csomag telepítése UHU-Linux alatt).
- kézikönyv: A The Not Too Short Introduction to LATEX2e c. könyv magyar fordítása (LATEX 78 percben címmel) letölthető a MagyarLATEX honlapjáról: http://www.math.bme.hu/latex/. Megjelent a LATEX kézikönyv.
- honlap, fórumok, levelezési listák stb.: a MaT_EX honlapon egy jó összefoglalás található az élő L^AT_EX-es projektekről



Magyar szöveg készítése LATEX-hel

Áttekintés

Honosítás (L10N)

Magyar szöveg készítése L^AT_EX-hel

Ki mit tesz MagyarLAT_EX Ábra Bábelben

Gyakori hibák

Összefoglalás

- ékezetes betűk: öt szép szűzlány őrült írót nyúz; betűtípusok
- a forrásfájl kódolása: inputenc, Latin-2, UTF-8
- *írásjelek*: "külső »belső« idézet" az ám!
- automatikus szóelválasztás: ékezetes betűk, dupla többjegyű mássalhangzók, összetett szavak, idegen szavak
- formai elemek: élőfej, lábjegyzet, fejezetszámozás, tételszámozás stb.
- nyelvi elemek generálása: a/az névelők, toldalékok; \az+\refstruc{xy+ban} → az 5. fejezetben
- tárgymutató: rendezési sorrendje és tipográfiája
- irodalomjegyzék: tipográfiája és rendezési sorrendje
- helyesírás-ellenőrzés: ISpell, MySpell, HunSpell

Az ezekhez szükséges szoftverek és leírások letölthetők a MagyarIATEX honlapról.



Ki mit tesz

Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel

Ki mit tesz

MagyarLATEX Ábra Bábelben

Gyakori hibák

Összefoglalás

- rendszergazda: telepíti a L^AT_EX-et és a magyar szöveg szedéséhez szükséges betűtípusokat és kiegészítő szoftvereket
- *tipográfus, szedő:* eldönti, hogy az adott műben mely szoftvereket milyen beállítással kell használni, meghatározza az egyes formai elemek tipográfiáját (pl. betűtípus, térköz)
- *szerző*: saját könyvtárába telepíti a még hiányzó szoftvereket; a kézi beavatkozást igénylő pontokon magyarítja a dokumentumot
- *szerkesztő, lektor*: ellenőrzik, hogy az elkészült mű megfelel-e a magyar tipográfiai szabályoknak
- *programozó*: L^AT_EX-makrókat és egyéb programokat a ír a mű egyedi problémáinak megoldására

Egyes szerepek feladatait gyakran ugyanaz a személy látja el.



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a L^AT_EX tölti be. Az eredeti magyar.ldf sürgősen felülírandó! \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xindy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a L^AT_EX tölti be. Az eredeti magyar.ldf sürgősen felülírandó! \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xindy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a I₄TĒX tölti be. Az eredeti magyar.ldf sürgősen felülírandó!
 \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- *husort.pl*: tárgymutató-feldolgozó, a *makeindex* helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xindy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a I₄TĒX tölti be. Az eredeti magyar.ldf sürgősen felülírandó!
 \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xindy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a I₄TĒX tölti be. Az eredeti magyar.ldf sürgősen felülírandó!
 \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xîndy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a I₄TĒX tölti be. Az eredeti magyar.ldf sürgősen felülírandó!
 \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xindy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez



Áttekintés

Honosítás (L10N) Magyar szöveg készítése L^AT_EX-hel Ki mit tesz

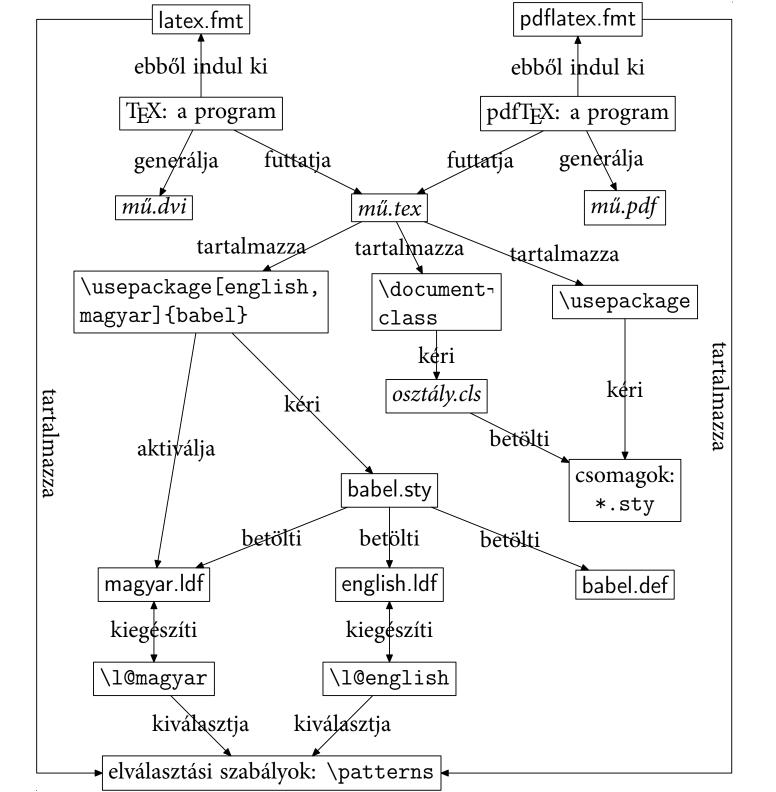
Magyarl^AT_EX

Ábra Bábelben

Gyakori hibák

Összefoglalás

- magyar.ldf (v1.5): Babel stílusfájl, a I₄TĒX tölti be. Az eredeti magyar.ldf sürgősen felülírandó!
 \def\magyarOptions{defaults=hu-min}
- huhyph*.tex: négy elválasztási mintafájl összegyűjtve
- husort.pl: tárgymutató-feldolgozó, a makeindex helyett
- huplain.bst: irodalomjegyzék-stílus, a plain.bst helyett
- ccs_extract.pl: dupla többjegyű msh-k elválasztását javítja
- *magyarp.xdy: Xîndy* stílusfájl magyar tárgymutatókhoz (tesztelése még hiányzik)
- *lafmtgen.pl*: T_EX formátumgeneráló szkript az elválasztási modulok könnyebb telepítéséhez





Ékezetes betűk

Áttekintés

Gyakori hibák

Ékezetes betűk

Idegen ékezetek Elválasztás Elválasztás (2) Szőrös betűk Írásjelek Képletek Formai elemek Formai elemek (2) Számozás

Összefoglalás

Az itt leírt hibák és megoldásaik megtalálhatók a MagyarI^AT_EX honlapon is.

Ha az ékezetes betűk nem jelennek meg vagy hibásak

A preambulumba írni: \usepackage[latin2]{inputenc}. Ha a forrásdokumentum nem Latin-2, hanem UTF-8 kódolású, akkor az *ucs* csomagot be kell tölteni, és az *inputenc*-nek utf8-at kell megadni. Szükség esetén az *ucs* csomag legfrissebb verzióját le kell tölteni a CTAN-ről. Egyes szövegszerkesztők 3 karakteres UTF-8 BOM-ot helyeznek a forrásfájl elejére. Ez nem működik LATEX-hel, BOM nélkül kell menteni (alternatív megoldás a LATEXkézikönyvben). Nincs beépített támogatás arra, hogy egy művön belül a forrásfájlok különböző bemeneti kódolásúak legyenek.

Az új *magyar.ldf* figyelmeztető üzenetet jelenít meg pl. az *inputenc*, *t1enc* csomagok hiánya esetén.



Idegen ékezetek

Áttekintés

Gyakori hibák

Ékezetes betűk

Idegen ékezetek

Elválasztás
Elválasztás (2)
Szőrös betűk
Írásjelek
Képletek
Formai elemek
Formai elemek (2)
Számozás

Összefoglalás

Ha hullámvonalas õ és kalapos û jelenik meg

- Ha \usepackage[latin2]{inputenc} van megadva, akkor ez nem fordulhat elő. (Ha *latin1* van megadva, akkor javítandó *latin2*-re.)
- UTF-8-as forrás esetén a billentyűzet-kosztást és a szövegszerkesztőt kell rávenni, hogy jó ékezetes betűket rakjon a forrásfájlba.
- Egy rosszul elmentett UTF-8 forrás egy alkalmas Perl szkripttel javítható, például:

```
perl -pi -e 'BEGIN{@H{"\xB5","\xBB","\x95","\x9B"}
=("\x91","\xB1","\x90","\xB0")}s@\xC3([\xB5\xBB
\x95\x9B])@\xC5$H{$1}@gx' forrás.tex
```

A fenti megoldás akkor is működik, ha a forrásfájl vegyesen tartalmaz jó és rossz ékezetes betűket.



Elválasztás

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek

Elválasztás

Elválasztás (2) Szőrös betűk Írásjelek Képletek Formai elemek Formai elemek (2) Számozás

Összefoglalás

Ha a LATEX rosszul választja el a szavakat, vagy nem is választ el

- Az elválasztási mintákat tartalmazó forrásfájlokat telepíteni kell. Ezek: huhyph3.tex, huhyphf.tex, huhyphc.tex és huhyphn.tex. Az egyiket át kell másolni huhyph.tex névre (alapból a huhyph3.tex szokott ez lenni).
- A mintafájlok egymás alternatívái, egyikük se tökéletes (az idegen és összetett szavak miatt: meg-int vagy me-gint?). A Huhyphn-ről lásd Nagy Bence előadását.
- A *language.dat* megfelelő sorait engedélyezni kell, például magyar huhyph.tex és magyarn huhyphn.tex.
- Újra kell generálni a T_EX formátumot. Ez vagy menüből, vagy fmtutil -all kiadásával lehetséges.
- A preambulumba: \usepackage[magyar]{babel}.



Elválasztás (2)

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás

Elválasztás (2)

Szőrös betűk Írásjelek Képletek Formai elemek Formai elemek (2) Számozás

Összefoglalás

Ha csak egy-két szó elválasztásával van baj

- \usepackage{t1enc} szükséges az ékezetes szavak automatikus elválasztásához.
- Ha a szóban van kötőjel, akkor a további automatikus elválasztáshoz a kötőjelet '--ra kell cserélni. Ezt a *magyar.ldf* biztosítja.
- A dupla többjegyű mássalhangzót a forrásban 'előzze meg: me'ccsel elválasztása mecs-csel lesz. Ezt is a *magyar.ldf* biztosítja.
- Rossz elválasztás esetén \hyphenation{szó} kivétel veendő fel. A toldalékolt alakokat is.
- Lehet kísérletezni az egyes mintafájlok csereberéjével. Ehhez a *magyar.ldf* hyphenation= opciója használható, például hyphenation=huhyphn.



Szőrös betűk

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás Elválasztás (2)

Szőrös betűk

Írásjelek Képletek Formai elemek Formai elemek (2) Számozás

Összefoglalás

Acrobat Readerben olvashatatlan a szöveg és lassan jelenik meg

- Ha a *times* csomag betöltésekor vagy a *t1enc* csomag kihagyásakor a jelenség megszűnik, akkor a probléma oka, hogy a T1 kódolású *Computer Modern* (CM) fontok (= EC fontok) csak raszteres változatban vannak telepítve.
- Az EC fontok vektoros (Type 1) változát a *cm-super* fontcso-mag tartalmazza. Régebbi T_EX-disztribúció esetén a CTAN-ről kell felrakni (50 MB!).
- dvips -Pcm-super opció szükséges a *cm-super* fontok aktiválásához. *pdflatex* esetén erre nincs szükség.
- A *Latin Modern* (LM) fontok: vektoros, T1 és más kódolású CM fontok. Jobb minőségűek és kisebbek a *cm-super* fontoknál, de kevesebb glyph-et és kevesebb méretet tartalmaznak. Újabb disztribúciókban már benne vannak, és a CTAN-ről is letölthetők. Betöltendő az *lmodern* csomag.



Írásjelek

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás Elválasztás (2) Szőrös betűk

Írásjelek

Képletek Formai elemek Formai elemek (2) Számozás

Összefoglalás

Egyes írásjelek rosszul vagy rossz térközzel jelennek meg

- \usepackage{t1enc} szükséges pl. a ,,vesszős'' idézés nyitó ligatúrájához.
- \textqq az új magyar.ldf-ben: \textqq{külső \textqq{belső} idézet} → "külső »belső« idézet".
- A számozott és számozatlan felsorolások címkéje állítható az új *magyar.ldf* labelenums= illetve labelitems= opcióival.
- Párbeszédbeli megszólalás az új *magyar.ldf*-beli \mond-dal.
- Speciális írásjelek az új magyar.ldf aktív karakterével: '': "magyar" nyitó; '': "angol" nyitó; '<francia: "francia» nyitó; meg'=megáll, meg'-megáll; meg'_intem; Kiss Előd'-Nagy Pál;nátrium'|klorid;kutatás'"+fejlesztés; ''--két mínusz.
- Kell kis köz az írásjel előtt? Igen! (új *magyar.ldf*)



Képletek

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás Elválasztás (2) Szőrös betűk Írásjelek

Képletek

Formai elemek Formai elemek (2) Számozás

Összefoglalás

Egyes matematikai jelek rosszul vagy rossz térközzel jelennek meg

- Tizedespontot vagy tizedesvesszőt tartalmazó számok az új magyar.ldf \MathReal parancsával hozhatók létre.
- Matematikai képletben {,} helyett sima vesszővel is írható tizedesvesszű az új magyar.ldf mathhucomma=yes opciója mellett.
- A fatoriálisjel utáni térköz mindig jó az új *magyar.ldf*-fel.
- Az új *magyar.ldf* az alap térközöket (pl. \medmuskip) a magyar hagyománynak megfelelően csökkenti.
- Az új magyar.ldf-fel a 1+2+3+4+5+6+7+8+9+10+11+12+14+15 < 999 képlet bármelyik jelnél megtörhető, és a jel a következő sorban ismétlődik.



Formai elemek

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás Elválasztás (2) Szőrös betűk Írásjelek Képletek

Formai elemek

Formai elemek (2) Számozás

Összefoglalás

Egyes formai elemekben a sorrend, a kisbetű–nagybetű vagy a pont helytelen

- Már a régi *magyar.ldf* is javította a címoldali dátum megjelenését, és próbálta javítani fejezetcímek és képaláírások és a tartalomjegyzék tipográfiáját (pl. *Ábra 1*: helyett *1*. *ábra*.).
- Az új *magyar.ldf* helyesen javítja a fentieket, plusz az élőfejet (ezt csak a *letter*, *article*, *book* és *report* osztályokban), a tételszerű környezetek számozását, a függelékek számozását, a felsorolásokat, a leíró listákat, a \paragraph-okat, a lábjegyzeteket.
- Az új *magyar.ldf* magyarítja a *varioref* csomag kimenetét.
- Tételszerű környezetek számozásakor pl. a *theorem* csomag mellett szükség lehet a \theoremstyle{magyar-plain} kiadására.



Formai elemek (2)

Áttekintés

Gyakori hibák Ékezetes betűk Idegen ékezetek Elválasztás Elválasztás (2) Szőrös betűk Írásjelek Képletek Formai elemek

Formai elemek (2)

Számozás

Összefoglalás

Egyes formai elemekben a sorrend, a kisbetű–nagybetű vagy a pont helytelen

- Az új *magyar.ldf* együttműködik a L^AT_EX szabványos és az AMS dokumentumosztályokkal, a *amsmath*, *theorem*, *ntheorem*, *amsthm*, *fancyhdr*, *endnotes*, *caption2*, *nath*, *fnpara* és *cjhebrew* csomagokkal, és magyarítja őket.
- Az új *magyar.ldf*-ben a lábjegyzetek számozása igen jól. testreszabható.
- Az új magyar.ldf a régivel kompatíbilis módban fut. A tipográfiailag helyes módhoz a \def\magyarOptions{defaults =hu-min} parancs kerüljön a preambulumba.
- Magyar irodalomjegyzék BibT_EX-hel: \bibliographystle{huplain}.
- Magyar tárgymutató: *husort.pl* vagy *magyarp.xdy*.



Számozás

Áttekintés

Gyakori hibák
Ékezetes betűk
Idegen ékezetek
Elválasztás
Elválasztás (2)
Szőrös betűk
Írásjelek
Képletek
Formai elemek
Formai elemek (2)

Számozás

Összefoglalás

Generált számok fix névelővel vagy toldalékkal körülvéve

- A \ref, \pageref és \cite parancsok automatikusan generált számokat (pl. fejezet-, oldal-, képlet-, hivatkozás-) illesztenek be a dokumentumba. A névelőt és a toldalékot is generálni kell. Helytelen: az (\ref{xy})-ból kiindulva.
- A névelőket már a régi *magyar.ldf* is képes volt generálni az \aref, \apageref, \acite, \az parancsokkal és ezek nagy kezdőbetűs változataival. Pl. \Aref({xy}) képletből kiindulva. Az új *magyar.ldf* stabilabban végzi ugyanezt.
- Az új *magyar.ldf* többtucat toldalékot is képes generálni, pl. \Atold({xy})+bol{} kiindulva.
- Az új *magyar.ldf* fejezetszintek neveit is képes generálni, pl. \az+\refstruc{öf+ben} generálhatja azt, hogy "az 1.3. szakaszban", de azt is, hogy "a 2. fejezetben".



Együttműködés

Áttekintés

Gyakori hibák

Összefoglalás

Együttműködés

Továbbfejlesztés?

- A régi *magyar.ldf* számos csomaggal összeakadt. Ennek két fő oka volt: az aktív karakter gyanútlan kezelése a Babelben és a nem elég körültekintéssel felüldefiniált parancsok.
- Az új *magyar.ldf* jól működik a főbb osztályokkal és csomagokkal, beleértve az alábbi nagy csomagokat is: *hyperref*, *graphicx*, *pdfscreen*, *powerdot*.
- Az új *magyar.ldf* óvatosan bánik az aktív karakterrel, felülbírálva a Babelt.
- Az új *magyar.ldf* szolgáltatásai opciók használatával egyenként testreszabhatók és egyenként kikapcsolhatók. Így egy esetleges jövőbeli összeakadási probléma elhárításához nem szükséges programozói tudás.
- A MagyarI^AT_EX egyéb szoftvereinél nem merültek fel hasonló összeakadási problémák.



Továbbfejlesztés?

Áttekintés

Gyakori hibák

Összefoglalás Együttműködés

Továbbfejlesztés?

- A helyesírás-ellenőrzés teljes átgonodlást igényel, beleértve a szövegszerkesztőkhöz való integrációt.
- A tárgymutató és az irodalomjegyzék generálásának fejlesztése vagy akár újraírása nem ütközik akadályba.
- Elvi szempontból jó automatikus szóelválasztáshoz igénybe kéne venni pl. a *Szószablya* projekt szótárral dolgozó motorját, leváltva a T_FX minta alapú modelljét.
- Hibajavítások, kisebb fejlesztések a magyar.ldf-en elképzelhetők, de kompromisszumok és "intelligens" keresztbe varázslások nélküli megoldás csak a teljes Babel (és esetleg a L^AT_EX-kernel egyes részeinek) újragondolásával születhet, ezt hasznos lenne az Ω projekthez kapcsolódva végezni.
- Áttervezendő a fontkezelés, a karakter–glyph kapcsolat, a szkriptnyelv és a honosítási interfészek.

