

A Lexicographers' Workbench online szótárszerkesztő rendszer

Lexicographers' Workbench (LWB) is an XML-based online Dictionary Writing System developed by MorphoLogic Ltd. The author supervised the design and development of LWB. The aim of this paper is to give an overview of how LWB supports the work of lexicographers, providing some insight into the history of electronic dictionary publishing in Hungary, the motivation behind the development of this system, and the advantages of using XML in the dictionary publishing industry in general.

1. Bevezető

A Lexicographers' Workbench (LWB) a MorphoLogic által fejlesztett és üzemeltetett XML-alapú online szótárszerkesztő webalkalmazás. A szerző irányította a szoftverfejlesztési munkákat és végzi jelenleg a rendszer karbantartását. Jelen írásban szeretnénk a hazai lexikográfia művelői számára bemutatni a rendszer keletkezésének történetét, valamint rövid áttekintést adni annak használatáról, érintve a szótárkészítés új digitális korszakának technológiai követelményeit.

2. Az LWB előzményei és motivációja

A MorphoLogic a hazai elektronikus szótárkiadás egyik nagy úttörője. MoBiDic és MoBiMouse elektronikus szótári rendszereink keletkezésük idején egyedülálló technikai megoldásaikkal nemzetközi elismertséget is szereztek. Több nagy szótárkiadóval együttműködve a magyar piacon ismert kétnyelvű szótárak jelentős részét jelentettük meg ezekben a keretrendszerekben.

A szótárak elektronikus formában történő kiadásának azonban előfeltétele, hogy azok tartalma valamilyen a számítógép által is jól értelmezhető, kereshető formában álljon rendelkezésre. A mai napig nem ritka, az első MoBiMouse-szótárak készítésének idején pedig általános gyakorlat volt, hogy a szótárak szövegszerkesztő programok segítségével íródtak. Ezek a kéziratok a nyomtatott kiadások céljaira megfelelőek voltak, de önmagukban alkalmatlanok voltak arra, hogy más, számítógépes felhasználás alapját képezzék.

Amikor az adatábrázolás módjáról kellett döntenünk, választásunk a nemzetközi lexikográfiában is általánosan használt XML-technológiákra esett. Az XML (Extensible Markup Language) a strukturált dokumentumok, és általánosabban tetszőleges adatszerkezet leírására kialakított formális nyelv. Rendkívül széles körben elterjedt nemzetközi szabvány, melyet a webes

technológiák, többek között a weblapok megjelenését leíró HTML nyelvet is kidolgozó W3C konzorcium gondoz. Az XML nyelv és a kapcsolódó technológiák részletes bemutatása messze túlmutat jelen cikk keretein, de feltételezzük, hogy az olvasók egy jelentős része eleve ismeri ezeket, illetve az alapok számtalan internetes forrásból megismerhetők.

A szótárak tartalmának XML formátumúvá alakítása nem egyszerű feladat. Erre a célra külön szoftvert fejlesztettünk, amely a Markup Wizard nevet kapta. A rendszer képes a kéziratban elhelyezett formázások (pl. dőlt vagy félkövér betűk, betűtípusok stb.), szeparátorok (vesszők, pontosvesszők, speciális jelek, jelentésszámozás stb.), és egyéb formai jellemzők alapján az adott szótár struktúráját leíró környezetfüggetlen nyelvtan segítségével a szöveges rtf formátumú forrást XML dokumentummá konvertálni. Sajnos ez a konverzió általában nem végezhető el teljesen automatikusan, mert tapasztalataink szerint a kézzel írt és formázott forrásdokumentumok jellemzően a szócikkek 3-5%-ában tartalmazznak valamilyen kisebb-nagyobb formai hibát. Ezek egyenkénti megtalálása és javítása igen időt rabló és ezért költséges tevékenység.

Ezeknek a korai tapasztalatoknak köszönhetően hamar egyértelművé vált, hogy az elektronikus és a nyomtatott szótárkiadás egyaránt profitálna abból, ha a szótárak a jövőben eleve XML formátumban készülhének, illetve a forrásaik ilyen módon lennének karbantartva, fejlesztve. Ennek nem csak az lenne az előnye, hogy az egyre inkább elengedhetetlen elektronikus, illetve online verziók gyorsabban és olcsóbban lennének előállíthatók, hanem az XML technológia által biztosított szigor a szótárak minőségét is kedvezően befolyásolná.

Lássuk kicsit részletesebben is, hogy milyen előnyökkel jár az XML-alapú szótárkészítés:

- Az XML-dokumentumokban élesen elválik az adattartalom és annak felszíni megjelenítése. Ennek köszönhetően egy XML-alapú szerkesztőben a szócikkíróknak nem kell foglalkozniuk az egyes szócikkelemek formázásával, minden figyelmüket a tartalomnak szentelhetik. Az XML-ből számítógépes konverzióval előállított végső megjelenési formákban, vagyis a nyomtatásra szánt kéziratban, vagy az online megjelenő HTML-ben a formázások, a szeparátorok, a számozások stb. a médium igényeinek megfelelően és automatikusan jelennek meg.
- Egy XML dokumentumban a benne előforduló szerkezeti elemek lehetséges helye szigorúan szabályozható, tehát a szerkesztési szabályzatban meghatározott szócikkszerkezettől nem lehet véletlenül sem eltérni.
- Az előállt adat már munka közben is jól strukturált, így jól kereshető, ami lehetőséget teremt a szerkesztési munkát segítő további szoftvereszközök használatára. Ezek lehetnek a munkafolyamatot ellenőrző egyszerű

kimutatások, vagy akár bonyolultabb, a szócikkek tartalmát vizsgáló hibakereső scriptek, melyek olyan szerkesztési elvek betartását is ellenőrizni tudják, amelyek nem fogalmazhatók meg egyszerűen szócikkszerkezeti megköthetként.

Ahhoz, hogy a szótárak XML-ként készülhessenek, természetesen szükség van olyan szerkesztőprogramra, ami ezt támogatja. Általános célú XML-dokumentumszerkesztőkből ma már viszonylag széles választék áll rendelkezésre, azonban ezek nem feltétlenül elégítik ki a szótárkészítők speciális igényeit. Ilyen specializált szoftverrendszerek, ún. *Dictionary Writing System*-ek is léteznek már, de ezek általában rendkívül költséges megoldások, illetve a nagy külföldi kiadók többnyire saját fejlesztésű zárt rendszerekkel dolgoznak. A MorphoLogic ezért saját szótárkiadó partnerei számára kifejlesztette az LWB első változatát.

Az LWB 1.0 csak Windows operációs rendszerek alatt futtatható asztali alkalmazás volt, amely a szerkesztett XML-dokumentumot is a felhasználó gépén tárolta. Az egyes szótárakhoz tartozó felhasználói felületek elkészítése külön-külön programozási munkát igényelt. Az LWB-nek ez az első korai változata kielégítette azt az alapvető igényt, hogy az új szótárak eleve XML-adatbázisként jöjjenek létre, de számos hátránya is volt. Mivel teljesen offline rendszer volt, semmilyen formában nem támogatta a több helyről, vagy a több felhasználó által végzett közös munkát. Egy-egy új szótárprojekt beindítása jelentős szoftverfejlesztést is igényelt a szótáranként eltérő szerkesztőfelületek miatt. A rendszert minden gépen külön kellett telepíteni, és egyes esetekben különleges szövegbeviteli módok támogatása miatt ez nem is volt egyszerű művelet. A következő lépés, amely ezeket a hátrányokat kiküszöböli, egy olyan webalkalmazás létrehozása volt, amelynek használatához csak egy böngészőre van szükség, és a felhasználói felületét a szótárak szerkezetét leíró, egyébként is szükséges segédfájlból dinamikusan állítja elő. Ez lett az LWB 2.0, jelen cikkünk fő témája.

3. Az LWB fő jellemzői és architektúrája

Az LWB 2.0 egy XML-alapú szótárszerkesztő webalkalmazás, amelyet a MorphoLogic üzemeltet saját szerverén. Az alábbiakban kifejtjük, hogy pontosan miben is áll, illetve mivel jár a rendszernek ez a három fő tulajdonsága, tehát hogy *webes, szótárszerkesztő és XML-alapú*.

3.1 A webalkalmazás előnyei

Mivel online szolgáltatásról van szó, az LWB-ben végzett munkához csak egy böngészőre van szükség, így gyakorlatilag bármilyen internetkapcsolattal rendelkező számítógépen használható. Az adatok a webalkalmazás mögötti adatbázisban tárolódnak. Ezzel megszűnik a szoftver és az adatok mozgásával,

frissítésével járó kellemetlenség. A rendszerhez, és azon belül egyes szótárakhoz egyidejűleg tetszőleges számú felhasználó férhet hozzá, nem okoz gondot a közös munka nagyobb munkaközösségek számára. Az adatokhoz való hozzáférés jól szabályozható, a rendszerbe természetesen felhasználónévvel és jelszóval lehet csak belépni, a kommunikáció a kliens és a szerver között titkosított csatornán keresztül történik. A rendszer üzemeltetéséhez nincs szükség saját informatikai szakemberre, mert azt a szolgáltató végzi. Az egyetlen hátránya ennek a felépítésnek az, hogy a munkához folyamatos internetkapcsolat kell, de napjainkban ez már nem szokott tényleges gondot okozni.

3.2 Szótárspecifikus funkciók

Az LWB nem általános célú XML-szerkesztő, hanem kimondottan a szótárkészítés sajátos igényeihez igazított rendszer. A szótári XML-eknek van néhány igen jellemző tulajdonsága, amely megnehezíti kezelésüket a szélesebb körben elterjedt általános szerkesztőkben.

Ellentétben a legtöbb alapvetően szöveges XML-dokumentummal, a szótári XML-ek nagyon erősen strukturáltak és hierarchikusak. Míg egy általános dokumentumban a szövegelemeket csupán itt-ott tarkítják a funkciójukat és/vagy formázásukat jelölő XML-elemek, illetve a dokumentumok alapvető szerkezetét (fejezetek, címek stb.) mutatja csak a címkézés, a szótárakban gyakran több az XML-címke, mint a szöveg. Ez a fő szöveget és az elemek hierarchiáját egyaránt áttekinthetetlenné teheti, a szerkesztést igencsak megnehezíti. Az LWB-ben, alkalmazkodva a szótárak jellemző szerkezetéhez, az általános szerkesztőkben megszokott címkés nézet helyett csak a szöveget jelenítjük meg, a szócikkszerkezetet pedig fastruktúráként ábrázoljuk.

További alapvető szótári sajátosság, hogy az adatnak van egy kiemelkedő fontosságú egysége, nevezetesen a szócikk, amelyet első sorban a címszóval azonosítunk. Általános célú rendszerekben jobb híján minden egyes szócikket külön fájlban kell tárolni, és külön kell megnyitni a szerkesztéshez. Ha egy szótár összes szócikkét egy nagy XML fájlban tároljuk, akkor nehézkessé válhat a szócikkek megtalálása, illetve az adat a mérete miatt akár kezelhetetlenné is válhat. Az LWB eleve azzal a szemlélettel készült, hogy az adat tárolásának és szerkesztésének egysége a szócikk, azok megtalálása a rendszer alapvető funkciója.

Az LWB főként kétnyelvű szótárakhoz készült (természetesen egynyelvű szótárakhoz éppolyan jól használható), amelyeknél visszatérő probléma a különböző nyelvű szövegrészek bevitele. Különösen nagy gond ez eltérő írásrendszert használó nyelvek esetén. Az LWB-ben beépített virtuális billentyűzet található, amelynek segítségével tetszőleges nyelv bármilyen karaktere könnyen bevihető, de ezen túlmenően az egyes beviteli mezők nyelve

automatikusan változik, tehát a felhasználónak nem is kell foglalkoznia a megfelelő nyelv kiválasztásával.

Nem kizárólag szótárakban fordul elő, de ott nagyon gyakori, hogy vannak bizonyos gyakran ismétlődő szövegelemek, amelyeket minden esetben azonos formában szeretnénk megadni. Ilyenek lehetnek többek között a lexikai minősítések vagy a szófaji meghatározások. Az LWB lehetővé teszi, hogy ezeket az elemeket ne kelljen kézzel begépelni, hanem egy zárt listából kiválasztva szűrjük be azokat. Hasonlóan könnyen megoldható különféle speciális szimbólumok beszúrása. A kezelőfelületet a 4. fejezetben fogjuk részletesen bemutatni.

3.3 Az alkalmazott XML-technológiák

Az LWB, mint azt többször is hangsúlyoztuk, XML-alapú rendszer, azonban a felhasználónak egyáltalán nem kell jártasnak lennie sem az XML nyelv, sem a kapcsolódó technológiák használatában. A felhasználói felület a mögöttes nyers XML adat technikai részleteit elrejtí a szótárszerkesztő elől, és csupán a szöveges tartalmat mutatja egy olyan könnyen kezelhető formában, amely a szócikk logikai szerkezetét tükrözi. Ezzel párhuzamosan a szócikk nyomtatott nézete is folyamatosan látható. Anélkül, hogy a jelen cikk keretein túlmenő szoftvertechnológiai részletekbe mennénk, mégis érdemes a teljesség kedvéért a felhasznált XML technológiákat és azok szerepét közérthetően bemutatni.

Az LWB-ben minden egyes szótárprojekt fő összetevői az *XML-ként tárolt adat*, az *XSD séma*, és az *XSL*, valamint az ezt kiegészítő *CSS stíluslap*.

3.3.1 XML

Annak előnyeiről, hogy a szótári adat XML-formátumban tárolódik, már volt szó, most azonban tekintsük át részletesebben is ezeket az előnyöket.

Az XML nagyon széles körben elterjedt nemzetközi szabvány, ezért az ilyen formátumú adatok kezeléséhez bőven állnak rendelkezésre akár ingyenesen elérhető eszközök is. Az ilyen adatok keresése, manipulálása, más rendszerekbe való integrálása hosszú távon biztosított. Az XML-adat jól hordozható különböző operációs rendszerek és platformok között.

Mivel az XML a lehető legrészletesebben kódolja a szócikkek szerkezetét, az ilyen módon tárolt adat rendkívül jól kereshető, ellenőrizhető. Az XML dokumentumokhoz definiálható szigorú szintaxis biztosítja, hogy a szócikkek szerkezete ne térjen el a szerkesztési szabályzatban lefektetett elvektől.

Az XML-ben az adat és annak megjelenítése elválí egymástól, a létrehozott szótári adathalmaz gyakorlatilag egy jól strukturált adatbázis, amely egyformán alkalmas számítógépes felhasználásra, nyomtatott kiadvány előállítására, online vagy egyéb elektronikus megjelenítéshez. A megjelenítést ezen túlmenően is tudjuk szabályozni, lehetőség van a szócikkekből könnyebben áttekinthető egyszerű nézetet, vagy pl. az összes példát tartalmazó teljes nézetet generálni. A

megjelenítés módja és részletessége hozzáigazítható a felhasználás módjához, illetve a felhasználó egyéni igényeihez. Ilyen módon tudta a már említett MoBiMouse rendszer a képernyőn a keresett szó felett megjelenő buborékban a szócikk leglényegesebb elemeit tartalmazó kivonatos formát megjeleníteni, míg az ablakos nézetben a teljes szócikket mutatni. Hasonlóan ugyanabból a mögöttes adatból készül a Magyar Nyelv Nagyszótárának online változatában a nyomtatott és a webes megjelenítés, amelyek között a felhasználó szabadon választhat.

A megjelenítés rugalmasságának köszönhetően lehetőség van arra is, hogy a szócikkhez kapcsolódó metaadatokat is az XML-ben tároljuk, segítve különféle projektmenedzsment funkciók egyszerű megvalósítását. Ilyen metaadatok lehetnek a szócikk szerzője, szerkesztője, a szócikk állapota, a szerkesztők megjegyzései. Ezeket az adatokat megjeleníthetjük akkor, amikor a szócikkeket szerkesztik vagy korrektúrázzák, míg a végfelhasználók előtt természetesen mindig rejtve maradnak.

3.3.2 XSD

Az XSD séma az XML-dokumentumunk nyelvtanát leíró segédfájl. Az olvasók egy része valószínűleg ismeri az XSD elődjét, a DTD-t. Az XSD szintén W3C szabvány, és hasonlóan a DTD-hez az XML-dokumentum szerkezetét definiálja, de számos olyan kiegészítő funkciója van, amely alkalmasabbá tette az LWB-ben való felhasználásra. Ezeknek a kiegészítő funkcióknak a segítségével az XSD-ben tudjuk tárolni a szótárak szerkesztőfelületét meghatározó összes információt.

A megengedett szerkezeteken, vagy ha úgy tetszik a szócikk szintaxisán kívül itt adjuk meg az egyes szócikkelemek „barátságos” nevét, ami végül a felhasználó képernyőjén megjelenik. Az XML-szabvány korlátozza, hogy az elemeknek milyen nevet adhatunk, és a tárhellyel is spórolhatunk, ha a példákat pl. a tömör és ékezeteket nem tartalmazó nevű <eg> elemekbe csomagoljuk. A grafikus felhasználói felületen viszont sokkal jobban mutat és hasznosabb is, ha az elemeknek beszédes nevet adunk, így az <eg> elemeket mutathatjuk „*Példa*”-ként.

Említettük már, hogy a szövegbeviteli mezők nyelve automatikusan változik. Az egyes elemekhez rendelt nyelvi beállítások is az XSD-ben adhatók meg.

Minden elemhez megadhatunk továbbá egy olyan listát, ami az ott gyakran ismétlődő szövegrészeket, jeleket tartalmazza, illetve meghatározhatjuk az adott elemekben alkalmazható ún. inline formázásokat. Ezeket a rendszer könnyen elérhető gyorsbeszűrőgombok formájában fogja megjeleníteni.

Szintén az XSD-ben definiáljuk a már korábban említett kötött tartalmú elemek lehetséges értékeit. Az ilyen elemek (jellemzően szófajok vagy minősítések) értékeit legördülő listából választhatjuk ki, ezáltal bevitelük gyorsabb lehet, és garantált az egységes használatuk.

Azt is meghatározhatjuk, hogy egy új elem beszúrása esetén annak milyen leszármazott elemei kerüljenek be a dokumentumba automatikusan. A szótárak szerkezetében a kötelező és a gyakran alkalmazott elemek nem feltétlenül esnek egybe. Elképzelhető például, hogy az az egység, ami az ekvivalenst és az ahhoz kapcsolódó kiegészítő elemeket foglalja össze, nem kötelezően tartalmazza magát az ekvivalenst, mert helyenként azt egy parafrázis pótolja. Így a szócikkszerkezetben az ekvivalens opcionális elem lesz, ennek ellenére egy felhasználóbarát rendszertől azt várjuk, hogy az ekvivalenst befoglaló elem beszúrásakor az ekvivalens beviteli mezője automatikusan jelenjen meg, hiszen az esetek túlnyomó többségében erre szükség van. Erre való ez a beállítás.

A szerkesztőben a különféle szöveges elemtípusokhoz különböző háttérszíneket rendelhetünk, ami nagyban megkönnyíti azok felismerését munka közben.

Mint láthatjuk, az XSD számos LWB-specifikus információt tartalmaz, de szerencsére az XSD szabvány megengedi, hogy ún. annotációk formájában ilyen kiegészítő adatokat tároljunk a sémában. Ennek köszönhetően az LWB-ben használt XSD teljesen szabványos, és a hozzá tartozó XML adattal együtt bármilyen más szoftverben is felhasználható.

3.3.3 XSL és CSS

Végül essen néhány szó a nyomtatott nézet megjelenítéséhez szükséges XSL és CSS stíluslapokról. Hasonlóan a korábbiakhoz, ezek is általánosan használt W3C szabványok. Az XSL-fájl azt a transzformációt definiálja, ami az XML-adatból a közvetlenül megjeleníthető HTML-t generálja. A CSS az így előállított HTML-ben alkalmazott stílusok konkrét megjelenítését szabályozza.

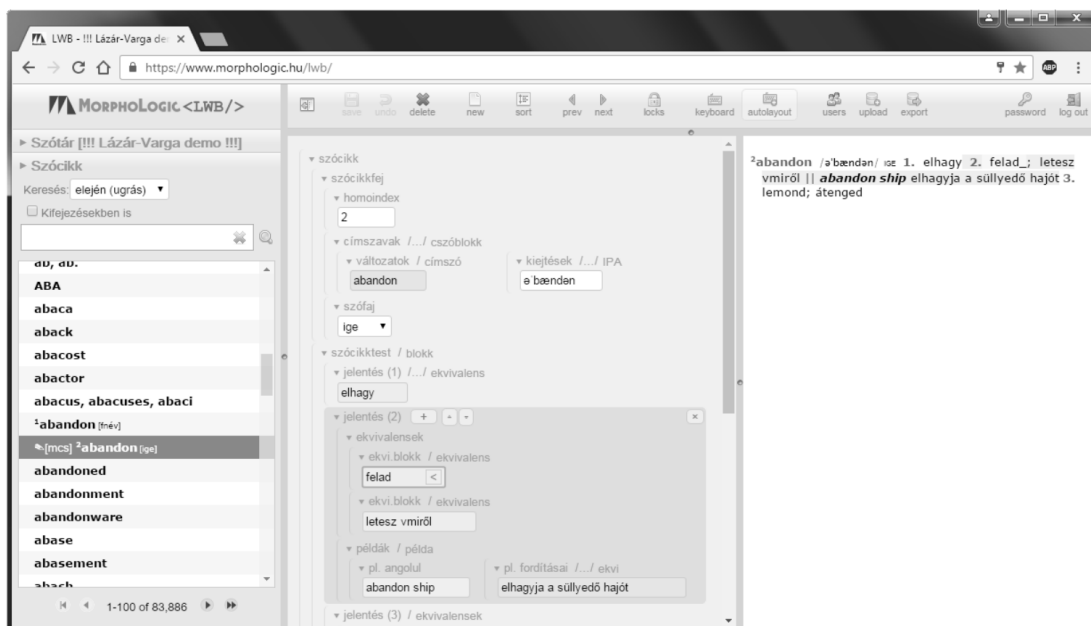
Ezekre a segédfájlokra egyrészt közvetlenül szükség van az LWB-ben az élő nyomtatott nézet előállításához, másrészt bármilyen további felhasználás alapját is képezhetik, legyen az nyomtatott kiadás, vagy online megjelenítés.

Általában a legegyszerűbb esetekben is az XSL felelős a szócikkekben a jelentésszámok és az egyes szócikkelemeket elválasztó szeparátorok automatikus elhelyezéseért. Az XSL-ben definiáljuk a szócikkelemek stílusát is, amelyek végső kinézetét a CSS-ben adhatjuk meg.

4. Az LWB funkcióinak ismertetése

Ebben a fejezetben az LWB funkcióit, illetve az azokat megvalósító felhasználói felületet ismertetjük képernyőképekkel gazdagon illusztrálva.

Az LWB-be felhasználónevünk és jelszavunk megadásával jelentkezhetünk be. A rendszer ezután felajánlja a felhasználó által elérhető szótárprojektek listáját. A szerkeszteni kívánt szótár kiválasztása után megjelenik az LWB szerkesztőfelülete (1. ábra).

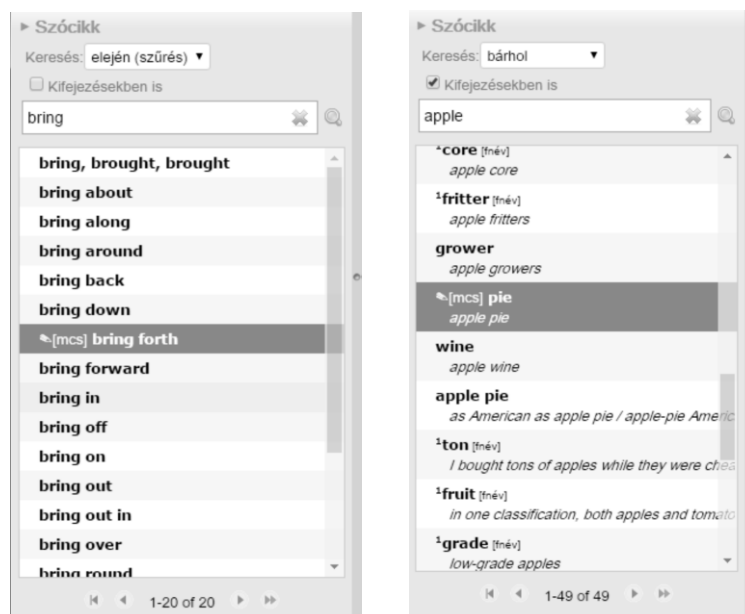


1. ábra: Az LWB szerkesztőfelülete

A szerkesztőfelület három fő részre oszlik. Bal oldalt található a *kereső*, középen a *szerkesztő*, jobb oldalon a szerkesztés alatt álló szócikk *élő nyomtatott nézete*.

4.1 Keresés

A keresőbe beírhatjuk a megtalálandó címszót, illetve megadhatjuk a keresés módját. Ennek függvényében a keresőmező alatt megjelenik a találatok listája. Az egyes találatokra kattintva betöltődik a szerkesztőbe a megfelelő szócikk (2. ábra).



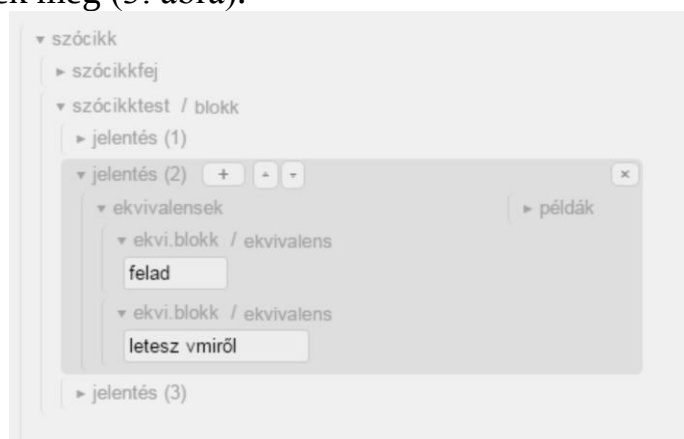
2. ábra: Keresési eredmények

Keresni lehet a címszavak elejére, végére, vagy tetszőleges részére. A megfelelő jelölőnégyzet kipipálásával lehetőség van arra, hogy a szótárban szereplő forrásnyelvi kifejezések tartalmára is rákeressünk. Eldönthető, hogy a találatok csak a feltételeknek megfelelő elemeket mutassák, vagy a teljes címszólistában ugorjunk az első találatra.

A találati lista mutatja azt is, hogy az adott szócikket pillanatnyilag szerkeszti-e valaki. A szerkesztés alatt álló, és emiatt átmenetileg zárolt szócikkek címszava előtt kis ceruzát ábrázoló ikon és zárójelben a szerkesztő felhasználóneve jelenik meg. Mások által szerkesztett szócikkeket nem nyithatunk meg. Amikor a szerkesztő kilép a szócikkből, annak zárolása automatikusan feloldódik, és elérhetővé válik más szerkesztők számára.

4.2 A szerkesztő

A szerkesztőfelületen minden XML-elemnek, vagyis a szócikkszerkezet minden elemének egy téglalap alakú terület felel meg. A faszerkezet levélelemeinek helyét elfoglaló szöveges elemek, amelyek a szótár tényleges szövegét tartalmazzák, ezeken a téglalapokon belül elhelyezkedő szövegbeviteli mezőkként jelennek meg (3. ábra).

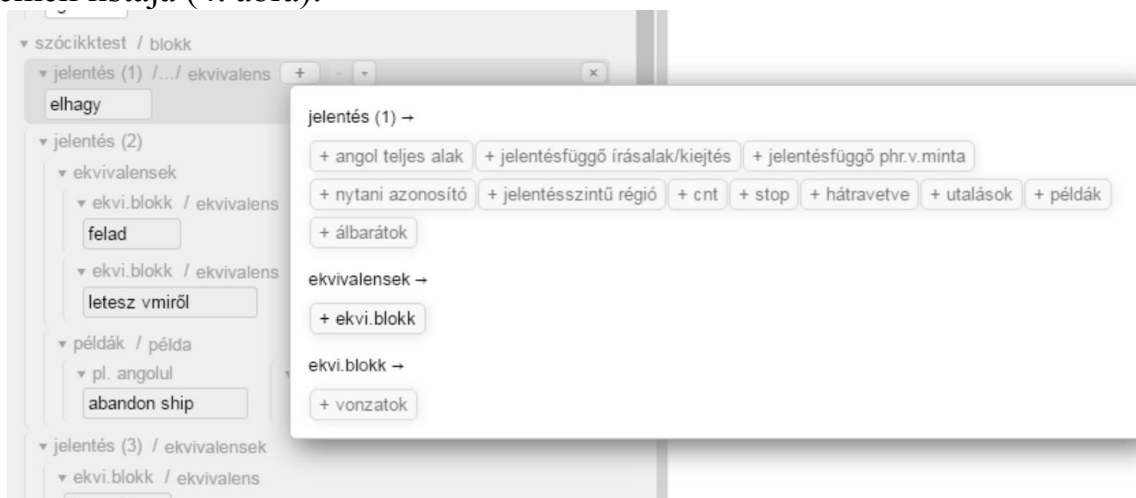


3. ábra: Elemek a szerkesztőfelületen

Az új elemek beszúrását, valamint az elemek mozgatását és törlését lehetővé tevő +, ▲, ▼, és × gombok az elem kijelölését, vagy az egérmutatónak az elem fölé mozgatását követően jelennek csak meg. Így a felhasználói felületet nem töltik meg ezek a zavaróan ismétlődő gombok, mégis egy kattintással elérhető minden alapvető funkció.

Az egyes elemek tartalmát befoglaló téglalap kinyitható és becsukható, a szócikkszerkezetet az elemek faszerkezetként való ábrázolása reprezentálja. A fenti 3. ábrán egyedül a második jelentés ekvivalensei vannak kinyitva, az összes többi szócikkelem be van zárva, így azok belső tartalma nem látható. Az elemek megmutatásához vagy elrejtéséhez az elem neve előtti kis háromszögre kell kattintani.

Új elem beszúrásához a szülőelem + gombja fölé kell mozgatni az egeret, aminek hatására megjelenik az adott pozícióban az XSD szerint beszúrható elemek listája (4. ábra):



4. ábra: Új elem beszúrása

A megfelelő gombra kattintva a szerkesztőben megjelenik az új elem, annak kötelező, valamint az XSD-ben meghatározott gyakran előforduló leszármazottjaival együtt.

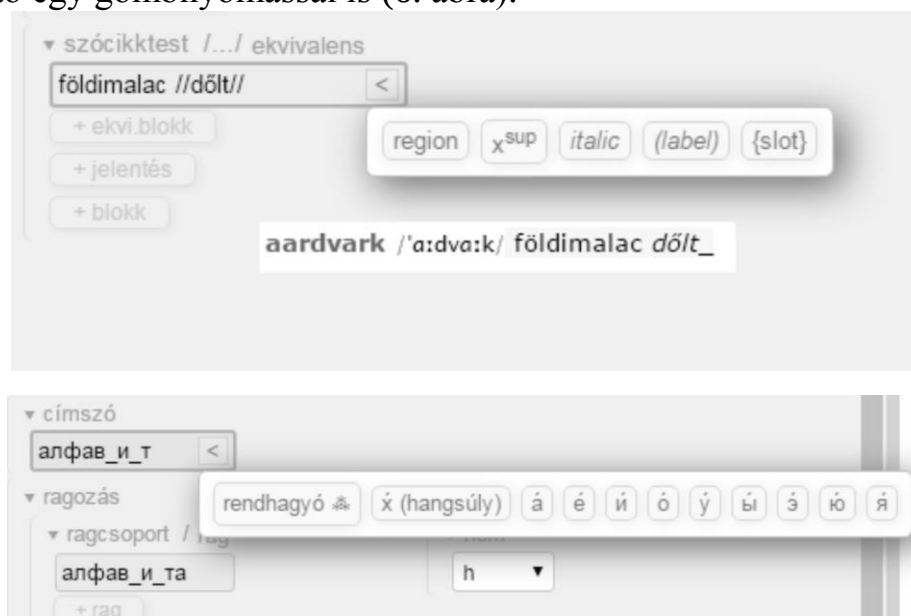
A szótárakban gyakoriak az ismétlődő elemek, mint például az egymást követő számozott jelentések, vagy az egy jelentéshez tartozó példák listája. Az ilyen ismétlődő elemekből egy új példány hozzáadása elvégezhető a lista alján megjelenő gomb segítségével is (5. ábra):



5. ábra: listaelemek beszúrása

A szövegbeviteli mezőkbe jobb esetben már egységes formázású szöveges tartalmak kerülnek, de a gyakorlatban szükségessé válhat az ilyen elemeken belül is különleges formázások alkalmazása. Ennek oka lehet egyszerűen a szöveg tartalma (pl. alsó, felső indexek), de ennél bonyolultabb is. Bizonyos szócikkelemek néha csak más elemek tartalmába ékelve jeleníthetők meg. Gyakori a rendhagyó igealakok ilyen módon való megjelölése, vagy a

hangsúlyok jelzése olyan esetekben is, amikor a nyelv helyesírása ezt nem teszi lehetővé. Néha szükségessé válhat glosszák szövegek közötti megjelenítése is, vagy összetételekben a fej nyelvtani nemének megadása. Az ilyen szövegek formázásokat ún. inline elemek beszúrásával valósítjuk meg. Az inline elemeket a beviteli mezőkben speciális kódok jelzik, de ezek beszúrása könnyen elvégezhető egy gombnyomással is (6. ábra):



6. ábra: Inline formázások és elemek

Az LWB más általános célú XML-szerkesztőkkel való összehasonlításakor már említettük, hogy a rendszer igyekszik a többnyelvű szövegbevitelt a felhasználó számára a lehető legkényelmesebb módon támogatni. Ennek érdekében integráltunk egy virtuális billentyűzetet, amelynek nyelve automatikusan változik az aktuális beviteli mezőtől függően. Az egyes mezőkhöz rendelt nyelv nem csupán létező természetes nyelv lehet, hanem tetszőlegesen definiált beviteli mód is. Így például könnyen megoldható a kiejtést tartalmazó elemeknél az IPA karakterek bevitelét segítő menü megjelenítése (7. ábra):



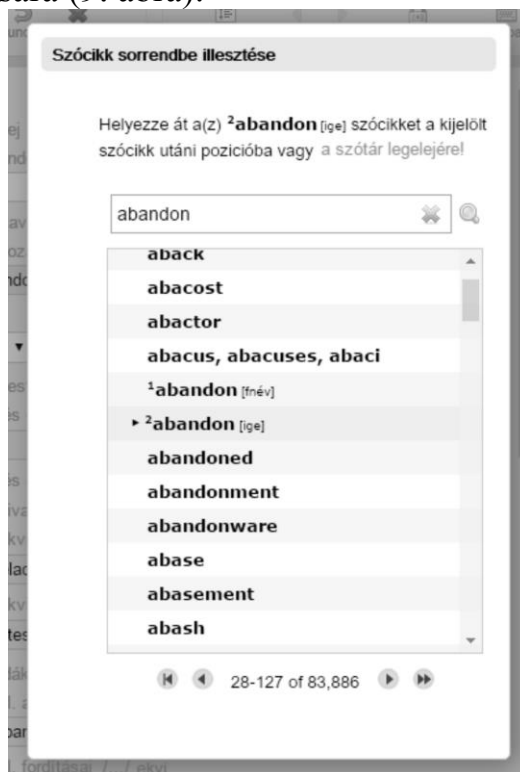
7. ábra: IPA bevitel

Ehhez hasonló módon működik pl. kínai írásjelek bevitele is (8. ábra):



8. ábra: LWB kínai szövegbevétel

Az LWB-ben a szócikkek sorrendje szabadon definiálható. Erre főként azért van szükség, mert a gépi rendezés nem minden esetben helyes. A magyarban nyelvi tudás, vagyis az összetételi határok ismerete nélkül a feladat nem is végezhető el automatizáltan. A rendszer új szócikkek létrehozása után elhelyezi azokat a gépi rendezés alapján adódó helyükre, de lehetőség van a szócikk helyének megváltoztatására (9. ábra):

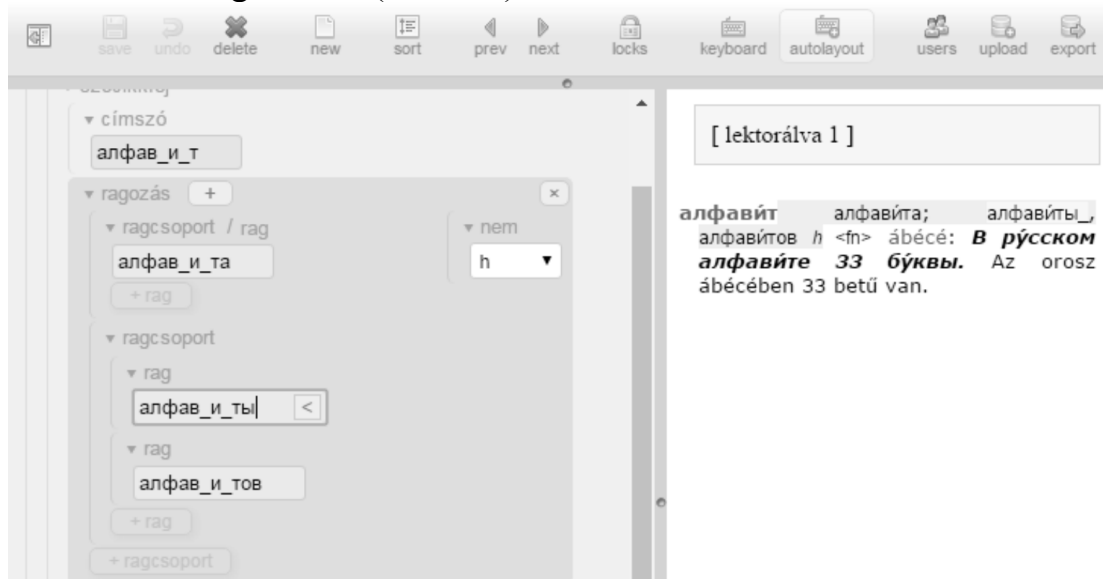


9. ábra: Szócikkek rendezése

A szótárak tartalmának XML fájlként való exportálása során a szócikkek az így meghatározott sorrendben kerülnek a kimenetbe.

4.3 Az élő nyomtatási nézet

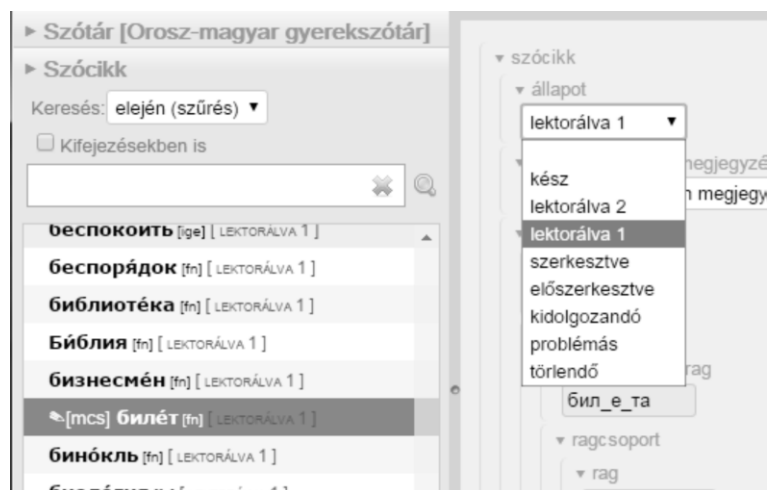
A szócikkek szerkesztése közben a képernyő jobb oldalán folyamatosan láthatjuk a végleges nyomtatási képet. A szerkesztőben kijelölt szerkezeti egység itt szürke háttérrel jelenik meg, az aktuálisan kijelölt beviteli mező tartalma pedig sárga háttérrel. Ez segíti a szócikk teljes tartalmának áttekintését, és az abban való eligazodást (10. ábra).



10. ábra: Nyomtatási kép kijelöléssel

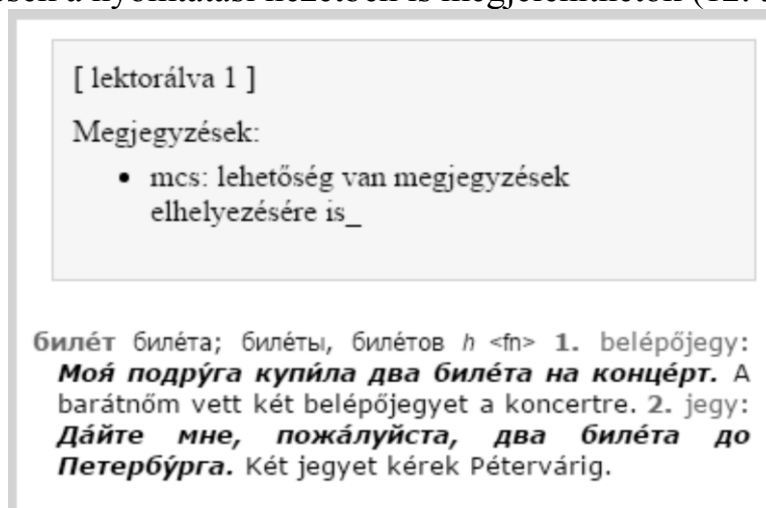
4.4 Projektszervezés

Az LWB-ben egyes egyszerű projektszervezési funkciók is megvalósíthatók. Ahogy azt már korábban is említettük, a szócikkekhez kapcsolódó metainformáció rögzíthető magában az XML-ben. Így megoldható például megjegyzések tárolása és megjelenítése, valamint a szócikkek állapotának nyilvántartása. A szócikkállapotok akár a címszólistában is megjelenhetnek, színkóddal segítve azonosításukat (11. ábra, az ábrán látható *LEKTORÁLVA 1* kódok az eredeti képernyőképen zöld színnel jelennek meg).



11. ábra: Szócikkállapotok

A megjegyzések a nyomtatási nézetben is megjeleníthetők (12. ábra):



12. ábra: Megjegyzések megjelenítése

5. Az LWB jelenlegi felhasználásai és jövője

A MorphoLogic által üzemeltetett éles LWB szolgáltatás jelenleg 75 szótárt kezel, amelyekben összesen 1.001.615 szócikk található. A Grimm kiadó majdnem teljes szótárportfólióját ebben a rendszerben tárolta és fejlesztette, egyes újabb kiadványok teljes egészükben az LWB-ben készültek. A MorphoLogic által kiadott, folyamatosan bővülő Lázár–Varga-féle angol–magyar elektronikus szótár is ebben a rendszerben íródik az elektronikus változat megjelenése óta. A havonta megjelenő állandó frissítések technikai alapja is az LWB.

Az LWB használhatóságát a fenti projektekben való sikeres alkalmazása és több éves megbízható működése bizonyítja, de természetesen ez a rendszer is fejleszthető. A legnyilvánvalóbb lehetséges fejlesztési irányok a keresési lehetőségek kibővítése, az XML attribútumok és a vagylagos szerkezetek

támogatása, valamint az online szótári rendszerekkel való közvetlen adatkapcsolat kiépítése lenne.

Legvégül említsük meg kiváló volt kollégáimat, Tőkés Tamást és Rác András, akik az LWB 1.0 és az LWB 2.0 rendszerek fejlesztését végezték. Külön köszönettel tartozunk Lázár A. Péternek és Varga Györgynek az LWB 2.0 fejlesztéséhez adott értékes tanácsaikért.