Csabay Károly
ELTE BTK Központi Könyvtár

*Dobó Juditnak**

Cédulák, ETO-szám, szócikkek

A cikk három egymástól viszonylag távol eső területről származó probléma bemutatására, és a közös megoldás ismertetésére vállalkozik. Formális nyelvi alapokról indulva végcélja katalóguscédulák, tezauruszok, szótárak adatelemeinek adatbázisba szervezése. Az elméleti apparátusnak és a ráépülő, a cikkben bemutatandó VIVALDI programnak a révén ez a cél elérhető közelségbe kerül. Teljes BNF-szintaxist közöl a könyvek bibliográfiai leírására, illetve ennek alkalmazását, egy konkrét bibliográfiai leírás teljes levezetési láncát adja meg. Hasonló szintaxist mutat be az ETO-számok formális nyelvi leírása céljából. Harmadik példája a szótári egység szintaxisa. A bemutatott problémákra közös megoldást nyújt a VIVALDI program: szkennelt katalóguscédulák retrospektív konverziója adatbázisimport segítségével; HTML – vagy bármilyen SGML-konverzió, vagy hasonló zárójelzés megoldása; egy bizonyos szintaxisnak megfelelő automatajellegű döntés; szócikkek és más tipografált formák analízise.

Noam Chomsky megállapítása szerint abban a lehetőségben rejlik a nyelv végtelensége, hogy a nyelv olyan struktúra, melynek eszköztára ugyan véges (szavak a szótárból, nyelvtani szabályok), de a potenciálisan elmondható mondatok száma végtelen. Ennek a végtelenségnek az igénybevételére matematikai értelemben akkor és csak akkor nyílik lehetőség, amikor a nyelvtani struktúrák úgy hivatkoznak egymásra, hogy egy korábban alkalmazott nyelvtani egység a kifejtés folyamán újra alkalmazhatóvá válik. A végtelenség tehát a *rekurzivitás*ban ragadható meg.

A következőkben formális nyelvi alapfogalmakat és néhány terminológiát használunk fel. Ezekben *Révész György* Bevezetés a formális nyelvek elméletébe (Budapest, Akadémiai, 1979.) c. művét követjük.

A bibliográfiai leírás

Korunk a retrospektív konverzió kora. A feladat az, hogy különböző papíralapú hordozókról mágneses hordozókra *„konvertáljuk"* könyvtáraink bibliográfiai tételeit. Nagyon leegyszerűsítve: a cédulákból rekordok keletkezzenek.

Ismeretes, hogy a bibliográfiai tétel – létrehozója szándékától függően – többféle lehet, nem tárgyaljuk ezért itt a bibliográfiai tétel kiegészítő adatait, például a raktári jelzetet, Cutter-számot, besorolási adatot, melléktételt és egyebeket. Kizárólag a könyvek bibliográfiai leírására szorítkozunk, és e téren is maradhatunk vitában az olvasóval. Reméljük azonban, hogy célunkat annyiban elérjük, hogy a módszer a bemutatottakból jól megismerhető, olyannyira, hogy könyvtáros és számítástechnikus kollégák együttműködve alkalmazni fogják és hasznát veszik.

A cél tehát még egyszer: a cédulákból rekordok keletkezzenek. A rekordokban pedig mezőket különböztetünk meg[1], *az tehát a feladat, hogy a bibliográfiai leírás adatelemei a funkciójuknak megfelelő mezőkbe kerüljenek.*

A kérdés tehát a következő: Honnan lehet tudni egy adatelemről (avagy még durvábban: egy betűsorról), hogy mi a funkciója? A szintaxis válasza: *A funkció a nemterminális.*

Tekintsünk mindjárt egy konkrét példát! Legyen a vizsgált bibliográfiai leírás a következő:

```
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről az iskola fennállásának 30., Bátaszék alapításának 851. évében. - Bátaszék : II. Géza Gimnázium, 1993. - 192 p. : ill. ; 20 cm
ISBN 963 04 3647 7 fűzött : ár nélkül
```

Kérdés, mi a funkciója ebben a leírásban a `Bátaszék` szó második előfordulásának? A válasz az, hogy ha sikerül találni egy levezetési láncot, amelyben Bátaszék nevének szóban forgó előfordulása (BNF formalizmust alkalmazva) egy

⟨Megjelenési hely⟩ ::= `Bátaszék`

helyettesítés alkalmazása révén bukkan föl, akkor a terminális elem (jelen esetben `Bátaszék`) funkciója a *Megjelenési hely* funkció, amit éppen ez a nemterminális mond meg.

---

* D. J. kedves tanárom, a formai feltárás tanára a debreceni KLTE-n. (csk)

Az *1. mellékletben* egy megoldást adunk a bibliográfiai leírás szintaxisára BNF formában. Ehhez két előzetes megfontolást kell tenni.

1. Nemcsak annak okán, mert a példában nem fér el annyi terminális elem, amennyire a gyakorlatban szükség lesz, hanem azért is, mert például a bibliográfiai leírások esetében a ⟨Megjelenési hely⟩ nemterminális helyére *bármely* település neve kerülhet az egész világról, sőt, esetleg később olyané is, amelyik egyelőre nem is létezik. Emiatt a BNF formulákban szerepeltetnünk kell majd olyan *kvázikijáratokat*, amelyek azt jelentik, hogy „itt bármi jöhet". Ezeket az absztrakt BNF metaszintaxis nem definiálta, a gyakorlatban mégis szükségesek.
2. Ismeretes, hogy a bibliográfiai leírás szabványa a sorok végének elhelyezkedéséről nem rendelkezik. A kikényszerített sorvég (hagyományosan két szóközzel a következő sor elején) a pont–gondolatjel kombinációval egyenértékű. A példa szintaxisa úgy tekinti a bibliográfiai leírást, mintha egyetlen (akármilyen hosszú) sorban helyezkedne el. A sorvég hovakerülésének kérdésével a továbbiakban sem foglalkozunk[2].

A BNF-ben szerepelnek kifejtetlen nemterminálisok (például ⟨**szöveg**⟩) ezeket félkövér betűkkel és kisbetűs írásmóddal különböztettük meg. Ezek közül némelyik (mint például a ⟨**pontosan tízjegyű arab szám**⟩) még kifejthető lett volna, de a BNF már így is elegendően hosszú. Ezeket a nemterminálisokat nem magyarázzuk tovább.

A *2. mellékletben* közöljük ebben a BNF-ben a bátaszéki II. Géza Gimnázium évkönyvéről készült bibliográfiai leírás teljes levezetését. Éltünk a lehetőséggel, hogy a BNF-ben tovább nem magyarázott nemterminális helyére egyszerűen behelyettesíthetjük az aktuális adatot.

A szintaxissal, illetve a levezetéssel kapcsolatban az alábbi észrevételeket tehetjük:

- Mivel a későbbiekben bemutatjuk a VIVALDI programot, mely a BNF szintaxist λ-mentes alakban várja, fel kell hívnunk a figyelmet, hogy az 1. mellékletben közölt BNF tartalmaz λ-kat. Ennek a BNF-nek is van λ-mentes ekvivalens alakja, de az – mint a λ-mentes nyelvtanok mindig – sokkal hosszabb az itt közöltnél; terjedelmi okból ezért ezt a változatot mutattuk be[3].
- Tennünk kell egy megkülönböztetést. Egy dolog az, hogy *a grammatikában mit lehet levezetni*, azaz hogy mi a grammatika által generált nyelv; és egy másik kérdés az, hogy egy konkrét elemző egy adott mondatot egy adott grammatika mellett *ténylegesen hogyan vezet le*. A nyelv absztrakt halmaz, amely az elemzőtől és a grammatika helyettesítési szabályainak felsorolásai sorrendjétől függetlenül létezik, és egyértelmű. A levezetés viszont esetleges – esetünkben például függhet a most közölt BNF sorainak sorrendjétől[4]. A bátaszéki évkönyv bibliográfiai leírásának a 2. mellékletben közölt levezetése nem az egyetlen lehetséges levezetés, még abban az értelemben sem, hogy egy másik levezetésben az alkalmazott helyettesítési szabályok *sorrendje* alkalmasint más lehet és lesz is. Olyan eltérő levezetés is adható, amely a célmondat[5] ténylegesen *más értelmezését* szolgáltatja. A leírásban a főcím tartalmaz pontokat. Tekintve, hogy a nyelvtan szerint az egyenrangú címeket és szerzőségi közléseket pontok választják el egymástól, a mi főcímünk elvileg akár öt egyenrangú részből is *állhat*[6]. Az emberi szemnek természetes, hogy a főcím pontjai ebben az esetben nem elválasztójelek, de nekünk ezt már a szöveg *szemantikája* mondja meg; a *szintaktikus elemző* nem dolgozik szemantikával. Ha szükséges, egyéb szintaktikai szempontot kell megadnunk a BNF segítségével, például a szóban forgó esetben azt, hogy a címnek nagybetűvel kell kezdődnie[7].

A most bemutatott probléma nem új a bibliográfiai tükröztetés elméletében. Az IFLA 1978 óta hatályos szabványa úgy rendelkezik, hogy az érintett adatelemben annak elválasztójeleit saját formájukban leírni nem szabad, tételesen azonban csak a *pont–szóköz–gondolatjel–szóköz*, a *szóköz–per–szóköz*, a *szóköz–egyenlő–szóköz*, a *szóköz–kettőspont–szóköz* és a *szóköz–pontosvessző–szóköz* karaktersorozatok leírásának tilalmát mondja ki[8]. Elvileg ugyanígy tilos lenne a *pont–szóköz* felvétele is, ez a tilalom azonban olvashatatlanná tenné a címeket, ezért a szabvány sem alkalmazza.

- A II. Géza-féle probléma inkább gyakorlati jellegű, abban áll, hogy vannak olyan címek, amelyeknek az esetében a tisztán szintaktikai szempontokra támaszkodó elemző más értelmezést ad – a formális nyelvi szempontból különben egyenrangú levezetések közül másikat –, mint mi. Ha akarnánk, a BNF finomabbá tételével avagy a cím másmilyen tükröztetésével akár ki is küszöbölhetnénk a problémát. Nem tesszük, mert nem érezzük annyira fontosnak. Vannak azonban ennek a BNF-nek *sui generis* halott ágai, melyekre az elemzés elvi okokból soha nem kerít sort. Ilyen például az egyéb címadat ága. Ennek oka az, hogy a szöveg egyéb címadatait ugyanúgy kettőspontok választanák el az alcímektől, mint ahogyan az alcímeket is kettőspontok választják el az őket megelőzőktől, alkalmasint a párhuzamos címtől, annak hiányában a főcímtől. Ha felbukkan tehát egy kettősponttal elválasztott szelet a szövegben, abból a BNF értelmében – tetszik vagy nem tetszik – alcím lesz, ha több ilyen

van, akkor abból több alcím lesz, egyszerűen azért, mert az alcímek *előbb jönnek*, mint az egyéb címadatok. Egyéb címadatot tehát az elemző azért nem fog adni, mert az alcím a szintaktikai azonosság miatt elviszi az összes ilyen szeletet. Egyszerűen nem létezik ésszerű ok, amellyel egy elemzőt rá lehetne beszélni arra, hogy valahányadiktól kezdve az ilyen szeleteket ne alcímnek, hanem egyéb címadatnak tekintse, tudniillik az egyik kettőspont pontosan olyan, mint a másik.

Ugyanígy járnak a további közreműködők pontosvesszővel elválasztott csoportjai a szerzők csoportjai *mögött*. A helyzet analóg: a szerzők minden ilyen csoportot „elvisznek", további közreműködőt az elemzés nem tud adni.

- Emlékeztet az előző helyzetre az alábbi, mégis mélyebb oka van, ezért külön említjük. Az egyéb címadatok és a további közreműködők problematikáját úgy lehetne jellemezni, hogy *aki előbb jön, az mindent visz*. Ebben a BNF-ben az alcímek is így járnak bizonyos értelemben, de ez a történet érdekesebb. Képzeljük el, hogy van a címben párhuzamos cím is. Szabvány nem mondja ki, de az intuíció azt sugallja, hogy ha a főcímnek vannak alcímei, akkor a párhuzamos címnek is kell legyen, mégpedig *ugyanannyi* alcíme[9]. Természetesen ugyanez a helyzet egynél több párhuzamos cím mellett is. Ez a követelmény Chomsky-féle generatív grammatikával valóban megfogalmazható, csakhogy a grammatika ténylegesen *környezetfüggő* lesz, BNF-fel tehát nem írható le. Ezt a dilemmát a most közölt BNF akként negligálja, hogy amikor egy párhuzamos cím fellép, az alcímek mind a párhuzamos cím alcímei lesznek. (Itt a párhuzamos cím alcíme *jön előbb*, mint a „rendes" alcím, ezért van, hogy mindent visz.)
- A *Megjelenés* adatcsoportjával kapcsolatos az a talán nem lényegtelen észrevétel, hogy míg a többi *pont–gondolatjel*lel bevezetett adatcsoport mindegyikénél megengedett a λ használata, a *Megjelenés* esetében nem; és itt a λ nem véletlenül maradt el. A bibliográfiai leírás hét nagy adatcsoportja közül a *Cím és szerzőségi közlés* mellett a *Megjelenés* még az, amely *soha nem lehet üres*.
- A közvetlen rekurzió (iteráció) alkalmazása több helyütt is megfigyelhető a grammatikában, legelső előfordulása a ⟨Címek és szerzőségi közlések⟩. Ennek a BNF-definíciónak a jelentése az, hogy a ⟨Címek és szerzőségi közlések⟩ nemterminális tetszőleges számú, de legkevesebb egy darab ⟨Cím és szerzőségi közlés⟩ nemterminálissal helyettesíthető, és ha több van, akkor a ⟨Cím és szerzőségi közlés⟩ nemterminálisokat pontok választják el egymástól. Van azonban a közölt grammatikában távolabbi rekurzió is, olyan, amelynek az esetében az iterációs kör több nemterminálison halad át.

  A mellékletek terjedelmének leírásánál a ⟨Terjedelem adata⟩ nemterminális gömbölyű zárójelekbe téve újra alkalmazható. Ez azt jelenti, hogy (a pont–gondolatjel bevezető jeltől eltekintve) a teljes bibliográfiai főadatcsoport megismételhető. Így elvileg az sincs kizárva, hogy melléklet mellékletének a terjedelmét írjuk le[10].
- A közölt BNF az ISBN-számot tagolatlannak tekinti (vö. az eredetileg közölt, illetve a levezetett írásmódot!).
- Nyilván a ⟨Kivitel⟩ értéke nem csak a közölt négy változat lehet, illusztrációnak azonban e négy elegendő.
- BNF az esetlegesen előforduló külföldi valutákkal nem számol.

Megadtunk egy teljes grammatikát BNF alakban a könyvek bibliográfiai leírására mint szintaktikai egységre. Bemutattuk egy viszonylag egyszerű konkrét bibliográfiai leírás teljes levezetési láncát. Azt ígértük, hogy a `Bátaszék` funkcióját a szintaxisban az a nemterminális fogja azonosítani, amelyből levezettük. Nos ez így is van (annak a közbevetésnek a figyelembevételével, amelyet a ⟨**szöveg**⟩-szerű nemterminálisokról mondtunk). Az az utolsó nemterminális, amelynek a helyére a `Bátaszék` terminális lépett, a ⟨Megjelenési hely⟩ volt. Az a részlet tehát a célmondatban, amely a `Bátaszék` jelsorozatban ölt testet, *funkcióját tekintve* megjelenési hely. Abban az esetben tehát, hogyha a célmondat elemeit (ha így tetszik: darabjait) egy adatbázisrekordba kívánjuk elhelyezni, a `Bátaszék` darabot a *„Megjelenési hely"* mezőnek kell értékül adnunk – föltéve, persze, hogy van a rekordban ilyen jelentésű mező.

Még egyszer: A BNF létének az a jelentősége, hogy a tagolatlan cédulaképből kiemelte a „Bátaszék" szót, és rámutatott, hogy „Megjelenési hely".

Hogyha az utolsó nemterminálissal együtt szemlélve is kétség maradna egy adatelem funkcióját illetőleg, akkor ismét azt az elvet kell segítségül hívnunk, hogy *a funkció a nemterminális*. Mostani példánkban a funkciójával értelmezett `Bátaszék` voltaképpen `Megjelenési hely = Bátaszék`ké vált[11]. Ha az utolsó nemterminálissal mint funkcióval címkézett terminális nem elegendő az egyértelmű azonosításhoz, akkor az utolsó előtti nemterminálist is igénybe kell venni (és így tovább). Egy szerzőségi közlés lehet például a főcím szerzőségi közlése, avagy a sorozatra vonatkozó szerzőségi közlés. Ez a megkülönböztetés fontos lehet, alkalmasint más-más mező lesz fenntartva a két funkció számára a képzeletbeli céladatbázisban. Ez a megkülönböztetés az utolsó

nemterminálisok szintjén nem jelenik meg, hanem „föl" kell menni egészen a mondatszimbólumig, hogy a különbség kiderüljön.

Az ETO-számok

Mostani demonstrációnk a bibliográfiai leíráson kívüli adatelemek egyik fontos területére kalauzol bennünket, ez pedig az ETO-számok szintaxisának területe. Tételezzük föl, hogy a feladat az, hogy *föl kell ismerni a szövegfájlban, hogy hol van(nak) az ETO-szám(ok)!*

Az a segédeszköz, amelyet erre a célra igénybe vehetünk, ismét a szintaxis, illetőleg a grammatika. Ahogyan a fentiekben a bibliográfiai leírásnak mint szintaktikai egységnek adtunk grammatikát, a *3. mellékletben* vázlatos javaslatot teszünk egy grammatikára BNF formátumban, amely a megengedett ETO-számokat generálja.

Némelyik nemterminálist nem fejtettünk ki tovább, ezeket félkövér szedés különbözteti meg; a grammatika tanulmányozása így is lehetséges.

Újból hangsúlyozzuk, hogy a grammatikák *szintaktikai* és *nem szemantikai* eszközök. Ezért nem szabad, hogy zavaró legyen az a helyzet, hogy ez a grammatika az alosztásokról visszautal a főtáblázati számra. Ez szemantikai szempontból nyilván nem állná meg a helyét, tisztán formailag tekintve azonban a nyelvtannak „igaza van".

A most mondottakkal egyidejűleg és nem azok ellenére mégis lehet szemantikai vonzata egy ilyen grammatikának. Ha egy ETO-szám levezetése során annak egy részletét csak a például ⟨Földrajzi alosztás⟩ nemterminálison keresztül lehetett levezetni, akkor az a részlet földrajzi alosztás. A helyzet ugyanaz, mint amit a bibliográfiai leírás BNF-jénél mondtunk: ha egy feldolgozó adatbáziskezelőnek van olyan adatmezeje, amelybe a földrajzi alosztásokat kell tenni, akkor ez után az elemzés után ezt megtehetjük.

A mélyebb tanulmányozást folytatók számára felhívjuk a figyelmet az ETO-szám nyelvtanának most közölt változatából eredő néhány következményre:

- E szerint a grammatika szerint nem helyesek a széles körben használt `929[ ...]` (például `929[ 894.511Csokonai]`) megoldások[12].
- Alfabetikus alosztás után – ez nem tűnik ki közvetlenül a grammatikából, de magától értetődik – szemantikai tartalommal már nem lehet fölvenni semmit. Nem hibás, de a jelzetalkotó szándékát nem éri el például a `681.3IBM-PC(0.062)` jelzet, hiszen az alfabetikus alosztás „mindent elvisz", azaz a `(0.062)`[13] jelentése is elvész. Nagyon hasonló a helyzet a csillaggal: mivel nem tudjuk, hogy az idegen jelzetrendszerben milyen karakterek fordulhatnak elő, ezért a csillagtól jobbra az egész jelzet kiesik a szemantikai látókörből.
- A `681.3IBM-PC(0.062)` jelzet abból a szempontból is jó példa, hogy kötőjel van az alfabetikus részben. Lehet olyan elemzőt készíteni, amelynek a prioritási rendszere szerint előbb a korlátozottan közös alosztásokat kell felismerni, ennek során a kötőjelre kell támaszkodni[14].

Szócikkek és egyéb szintaktikus egységek

A harmadik modell, amely BNF formátumú nyelvtanok támogatását kívánja, egy szótár szócikkeinek elemzése. Elsősorban terjedelmi oka van, hogy itt további BNF-et nem közlünk, másodsorban azért nem, mert a most bemutatott feladatok megoldását jelentő számítógépes programot kívánjuk ismertetni.

Ehelyütt most arra utalunk, hogy a rendelkezésünkre álló hatalmas nyomtatott információs halmazban, melyet előszeretettel neveznek Gutenberg-galaxisnak is, óriási mennyiségben áll rendelkezésre valamilyen *tipográfiai, központozási*, esetleg *elhelyezési* úton (is) rögzített információ. Ismeretes, hogy adatszerkezet nélkül a parttalanul ömlő adat nagyon keveset ér, ha éppen nem semmit. Az adatszerkezet a most mondott jellemzőkben rejlik. Bibliográfiák, katalóguscédulák, tezauruszok, lexikonok, szótárak válhatnak hozzáférhetővé a számítógépes feldolgozás számára – reményeink szerint egy egyszerű szkennelés után – a grammatikák módszerének alkalmazásával.

A VIVALDI

A VIVALDI program, amely szintaktikus elemzést végez[15]. A program leírását a *4. mellékletben* adjuk közre. Elöljáróban emlékeztetjük olvasóinkat arra, hogy a BNF-ben szerepelhetnek kifejtetlen nemterminálisok (például ⟨**szöveg**⟩). Erre a célra a VIVALDI néhány beépített lehetőséget eleve nyújt.

Mire jó a VIVALDI? – a korábban fölvetett problémák

Adatbázisimport

Ha a VIVALDI-t fel akarjuk használni egy konkrét alkalmazás során, először is *el kell képzelni, hogy mit akarunk vele csinálni*, ennek megfelelően meg kell alkotni az alkalmazás BNF-jét; a BNF –

majdnem biztosan – nem λ-mentes, ezért λ-mentesíteni kell. A λ-mentesítés fárasztó és elhibázható mechanikus feladat, célszerű tehát programra bízni[16].

Illusztrációképpen bemutatjuk a *Könyvek bibliográfiai leírása* BNF-jének első fejezetét λ-mentesítés előtt és után. Emlékezetes; a fejezet így nézett ki:

⟨Könyvek bibliográfiai leírása⟩ ::= ⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩

A fejezet a λ-mentesítés után így alakul:

Könyvek bibliográfiai leírása::=
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Kiadás⟩ ⟨Megjelenés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Sorozat⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjedelem⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Sorozat⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Megjegyzés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Megjegyzés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩ ⟨Terjesztés⟩|
⟨Címek és szerzőségi közlések⟩ ⟨Megjelenés⟩

Amint látható, a BNF terjedelme igen-igen megnövekszik az eljárás következtében. Emellett még a *Könyvek bibliográfiai leírása* BNF-jének azokat a nemterminálisait is definiálni kell, amelyeket a korábbi tárgyalás során nem fejtettünk ki (emlékszünk például a ⟨**szöveg**⟩ nemterminálisra).

Egy így kiegészített BNF már működőképes a VIVALDI programmal. Tételezzük fel, hogy a λ-mentesített BNF a BIBL.BN4[17] állományban, a bátaszéki II. Géza Gimnázium évkönyvéről szóló bibliográfiai leírás pedig egyszerű ASCII szövegfájl formájában a BIBL.TXT állományban van, mégpedig az alábbi sortördelés szerint:

```
A bátaszéki II. Géza Gimnázium jubileumi
évkönyve az 1992-93. iskolai évről az
iskola fennállásának 30., Bátaszék
alapításának 851. évében. - Bátaszék :
II. Géza Gimnázium, 1993. - 192 p. : ill.
; 20 cm. -
ISBN 9630436477 fűzött : ár nélkül
```
[18]

Ekkor a VIVALDI egy

```
vivaldi /z bibl.bn4 bibl.txt bibl.out
bibl.hib
```

hívásával ellenőrizni lehet a BIBL.TXT (tehát a bibliográfiai leírás) szintaktikai helyességét.

Már ez is figyelemre méltó eredmény, de ekkor még a VIVALDI *„nem csinál semmit"* abban az értelemben, hogy csak ellenőrzést végez és transzformációt nem. A BIBL.OUT állomány ilyenkor csak visszatükrözi az inputot[19].

Abban az esetben, ha a VIVALDI-val átalakítást, transzformációt akarunk végeztetni, el kell döntenünk, hogy melyik nemterminális szinthez tartozó tartalmat kívánjuk az output állományba juttatni. Ennek megfelelően kell az alábbi jellegű kiegészítéseket elhelyezni a BNF-ben:

```
⟨Cím⟩ = $CIM: = $ ::=
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Alcímek⟩
        ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Alcímek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Alcímek⟩|
    ⟨Főcímek⟩ ⟨Alcímek⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩|
    ⟨Főcímek⟩
```

A fenti kiegészítés jelentése kettős:

- A BNF utasítja a VIVALDI-t, hogy amikor sikeresen elemzett egy ⟨Cím⟩ nemterminálist, akkor a terminális jelsorozat azon részlete elé, amely ebből a nemterminálisból származott, az outputban a CIM: címkét helyezze el.
- A dollárjelek nem látható, hanem formázó részletei az outputnak, a VIVALDI

```
vivaldi /l$ /p /z bibl.bn4 bibl.txt
bibl.out bibl.hib
```

hívásával utasítjuk a programot arra, hogy a dollárjeleket soremelésre cserélje[20]. A cím címkéjéhez hasonló további címkék elhelyezése után a már megismert inputból az például az alábbi output nyerhető:

```
CIM:A bátaszéki II. Géza Gimnázium
jubileumi évkönyve az 1992-93.
iskolai ...
  . -
  MEGJ.HELY:Bátaszék
   :
  KIADO:II. Géza Gimnázium
  , 1993. - 192 p. : ill. ; 20 cm. -
  ISBN:ISBN 9630436477
  fűzött : ár nélkül
```

A CIM: után ott áll a teljes cím, csakhogy soremelés nélkül, ezt jelzi az, hogy „lemegy" a lapról; a címkézett (értékes) outputok között ott látható a most érdektelen „hulladék" is[21].

Bizonyára sokan ismerünk olyan adatátalakító programokat, illetve adatbázisimport előtétmodulokat, amelyek a fenti adatformátumot – az úgynevezett *címkézett szövegfájl* formátumot – adatbázisba konvertálni képesek. Emlékeztetőül bemutatjuk a közismert DIALOG egy címkézett rekordját:

```
FN- DIALOG DOE Energy file 103
AN- 1178799
AN- <DOE> EDB-86:133382|
TI- <Analytic> Burundi peat project|
TI- <Monographic> Tropical peat resources
    - prospects and potential|
AU- Kalmari, A.;  Leino, P.|
CT- Symposium on tropical peat resources
    - prospects and potential|
CL- Kingston, Jamaica|
CY- 25 Feb 1985|
PU- International Peat Society,Helsinki,
    Finland|
PY- 1985|
PG- 340-349|
RN- CONF-8502128-  |
CP- Finland|
GL- Finland|
LA- English|
JA- EDB8606|
DT- Analytic of a Book; Conference
    literature|
AB-    Since 1980 EKONO Consulting
    Engineers, Finland has been working
    with a large peat project in Burundi,
    Africa. The work is being financed by
    Finnida, IDA, UNDP and the government
    of Burundi. ONATOUR from the Ministry
    of Public Works and Mines has been the
    local counterpart. Part I, completed
    1981, was a comprehensive study
    including resource survey,assessment
    of wet and dry harvesting methods, and
    processing of peat for a large nickel
    refining process. Phase II is almost
    completed and consists of large scale
    trial production using wet peat mining
    method combined with conventional
    milled peat and sod peat harvesting.
    The actual field work with weather
    monitoring, environmental data
    collection and trial area preparation
    started in 1981 and the actual
    production tests with full scale
    machinery have continued for almost
    one year. The trial project is
    scheduled to be completed by mid 1985.
    The paper gives an overview of the
    whole project and describes the EKONO
    underwater peat mining and production
    method developed for undrainable
    tropical peatlands.|
DE- <Major> áBURUNDI - COAL DEPOSITS;
    áBURUNDI - SURFACE MINING; áCOAL
    DEPOSITS - RESOURCE ASSESSMENT; áPEAT
    - PRODUCTION; áSURFACE MINING -
    PERFORMANCE TESTING|
DE-  COAL PREPARATION; DEMONSTRATION
    PROGRAMS; FIELD TESTS; FUEL
    SUBSTITUTION; HARVESTING; HYDRAULIC
    MINING; INTERNATIONAL COOPERATION;
    METAL INDUSTRY; NICKEL; RESOURCE
    DEVELOPMENT; RESOURCES; SURVEYS;
    UNDERWATER OPERATIONS|
DE- <Broader Terms> AFRICA; COOPERATION;
    DEVELOPING COUNTRIES; ELEMENTS; ENERGY
    SOURCES; FOSSIL FUELS; FUELS; GEOLOGIC
    DEPOSITS; INDUSTRY; METALS; MINERAL
    RESOURCES; MINING; ORGANIC MATTER;
    RESOURCES; TESTING; TRANSITION
    ELEMENTS|
SC- 011000á; 012000 ||
```

A VIVALDI-val kapcsolatos munkahipotézis tehát a következő:

1. A katalóguscédula szkennerrel beolvasható szövegfájlba.
2. Ez a fájl nagyon egyszerű automatikus előfeldolgozás (például a bekezdések helyettesítése pont-gondolatjelekkel) után VIVALDI-inputtá tehető.
3. A VIVALDI-output vagy közvetlenül, vagy egy automatizálható lépés után adatbázisba importálható.

HTML-konverzió

Az eddig elmondottakból már kitalálható, de jelentősége miatt külön érdemes kitérni arra a lehetőségre, hogy a VIVALDI a sikeresen elemzett terminálissorozatot nemcsak címkézni, hanem zárójelezni is képes. Ha a BNF-ben nem a fönt említett módosítást, hanem például a

```
⟨Cím⟩ = ⟨H1⟩ = ⟨/H1⟩ ::=
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Alcímek⟩
        ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Alcímek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Alcímek⟩|
    ⟨Főcímek⟩ ⟨Alcímek⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Egyéb címadatok⟩|
    ⟨Főcímek⟩ ⟨Párhuzamos címek⟩|
    ⟨Főcímek⟩
```

módosítást hajtjuk végre, akkor a címet az outputban a ⟨H1⟩, ⟨/H1⟩ zárójelek között találjuk. Ugyanígy lehet megadni bármi más SGML parancspárt egy környezetfüggetlen grammatikával azonosítható szövegrészen, és hasonlóképpen ruházhatjuk fel tipográfiai előkészítés céljával szövegeinket Word RTF vagy TEX-parancsokkal (illetve zárójelekkel).

ETO-számok felismerése

Természetesen van arra mód, hogy a VIVALDI megfelelő interfész kialakítása után más programokból hívható rutin legyen. Ekkor egy korábban rögzített BNF mellett tetszőleges neki átadott szöveg szintaktikus elemzését elvégezheti a főprogram számára. Az ETO-számok felismerésével kapcsolatosan ismertetett probléma például egy ilyen összetett alkalmazással kezelhető. A feltételezett VIVALDI-rutinnak el kell döntenie, hogy a vizsgált katalóguscédula-sor ETO-szám-e. Ebben a példában a VIVALDI-tól nem várunk adatátalakítást, csak egy igen/nem választ.

Szócikkek elemzése

Befejezésül egy igen elegáns alkalmazást: szótári szócikkek elemzését mutatjuk be.

A kiindulási feladat a következő: Adva van egy szótár – adott esetben egy „rendes" kétnyelvű szótár –, amelynek szócikkei szkennelés után szövegfájl formában rendelkezésre állnak. Célunk elkészíteni a szótárnak abban az értelemben vett inverzét, amely lehetőséget ad a jobb oldalon álló minden nyelvi elemnek fordított irányú keresésére. Példa:

*s***Leid** [~(e)s] **1.** fájdalom; bánat; sérelem, sértés; **sich ein ~(s) antun** kárt tesz magában; **sein ~ klagen** *a)* panaszkodik; *b)* elpanaszolja, hogy milyen sérelem érte; *c)* baját/bánatát (el)panaszolja; **jm ein ~ zufügen/tun** *a)* vkinek bajt/fájdalmat okoz; *b)* vkit (meg)bánt **2.** panasz, sirám **3.** töredelem, megbánás

Ebből a szócikkből, amikor a magyar oldalról keresünk, a jobb oldalon fellelhető minden nyelvi elemre kell tudnunk keresni; így nemcsak a *fájdalom*, *bánat* stb. szavakra, hanem a *kárt tesz magában* kifejezésre is.

A VIVALDI alkalmazásával létrejött megoldás alapgondolata az, hogy minden szótári szócikk egy egységes, strukturált formátumot követ. Ez a formátum BNF-fel megadható. Ezután a VIVALDI a szócikkekből szintaktikai helyüknek megfelelően fölcímkézett outputot állít elő, mely output (szükség esetén) adatbázisba importálható, illetve például a MorphoLogic MoBiDic programjával MoBiDic-szótárba szervezhető[22].

Szócikkek szerkezeti felépítésének illusztrálására mutatjuk be a következő struktúrát, amelyet *Tihanyi László*, a MorphoLogic munkatársa bocsátott rendelkezésre.

Az iskolai szótár szócikkéről készült BNF egy részlete:

```
# ISI.BNF (c) Tihanyi '97.jul.10
#1
<entry>=<entry>#=</entry>#::=
 <form><body>
#2
<form>=<form>#=</form>#::=
 <orthPart><formlabel>
#2a
<formlabel>::=
<pastform><pron><gramD><usgD>|
<pastform><pron><gramD>|
<pastform><pron><usgD>|
<pastform><pron>|
<pron><gramD><usgD>|
<pron><gramD>|
<pron><usgD>|
<pron>
#3
<body>::=
<xref>|
<homMultis>|
<homMono>
```

A közölt részlet segítségével megfigyelhetjük a # karakter felhasználását megjegyzések elhelyezésére, valamint hogy például az ⟨entry⟩ értékű részt Tihanyi instrukciója szerint ⟨entry⟩–⟨/entry⟩ zárójelpárba kell helyezni az outputban[23].

Végezetül bemutatjuk a Leid szócikkből képződött outputot[24]:

```
CIMSZO=»Leid«
NEM= n
NYELVTAN= [ =(e)s]
ARAB= »1.«
MAGYAR= fájdalom
MAGYAR= bánat
MAGYAR= sérelem, sértés
WENDUNG=»sich ein ~(s) antun«
MAGYAR= kárt tesz magában
WENDUNG=»sein ~ klagen«
MAGYAR= panaszkodik
MAGYAR= elpanaszolja, hogy milyen sérelem
érte
MAGYAR= baját/bánatát (el)panaszolja
WENDUNG=»jm ein ~ zufügen/tun«
MAGYAR= vkinek bajt/fájdalmat okoz
MAGYAR= vkit (meg)bánt
ARAB= »2.«
MAGYAR= panasz, sirám
ARAB= »3.«
MAGYAR= töredelem, megbánás
```

További feladatok – párhuzamos utak

A VIVALDI képességei bizonyos pontokon korlátozottak. Nem szóltunk még róla, de egyelőre van még egy paraméter, amellyel nagyon körültekintően kell számolni a VIVALDI alkalmazásakor, ez pedig az idő. A program demonstrációra – például egyetemi oktatási célra is – kiválóan alkalmas, de 60–70 ezer címszavas kéziszótár teljes anyagának elemzését csak kiemelkedően jó teljesítményű[25] PC-vel ajánlatos elindítani, és még ekkor is számolni kell azzal, hogy éjszakára magára kell hagyni a számítógépet.

Komoly fejlesztésnek tehát csak akkor van realitása, hogyha a programot sikerül megfelelő[26] fejlesztőkörnyezetbe áttelepíteni. Ezen a téren azonban mértékadó tapasztalatok még nincsenek.

A VIVALDI-éhoz hasonló teljesítményt nyújthat (az elsősorban UNIX környezetben ismert) megfelelően kialakított *yacc*, illetve *lex* alkalmazás. Tihanyi mérései ezen a területen igen kedvező időadatokat mutatnak. Tudni kell azonban, hogy ezek az eszközök további megszorításokat követelnek meg: a BNF által megadott környezetfüggetlen nyelvtannak például LR(1) grammatikának kell lennie. Ez igen erős megszorítás, ezzel szemben a VIVALDI *lényegesen nem egyértelmű grammatikák* kezelésére is képes[27]. A szerző minden ilyen irányú tapasztalatot, mérést, illetve javaslatot szívesen fogad[28].

1. melléket
Könyvek bibliográfiai leírása – a grammatika

⟨Könyvek bibliográfiai leírása⟩ ::=⟨Címek és szerzőségi közlések⟩⟨Kiadás⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Sorozat⟩⟨Megjegyzés⟩⟨Terjesztés⟩

⟨Címek és szerzőségi közlések⟩ ::=⟨Cím és szerzőségi közlés⟩. ⟨Címek és szerzőségi közlések⟩|⟨Cím és szerzőségi közlés⟩

⟨Cím és szerzőségi közlés⟩ ::=⟨Cím⟩⟨Szerzőségi közlés⟩

⟨Cím⟩ ::=⟨Főcímek⟩⟨Párhuzamos címek⟩⟨Alcímek⟩⟨Egyéb címadatok⟩

⟨Főcímek⟩ ::=⟨Főcím⟩ ; ⟨Főcímek⟩|⟨Főcím⟩

⟨Főcím⟩ ::=⟨Főcím adata⟩⟨Alcímek⟩⟨Szerzőségi közlés⟩

⟨Főcím adata⟩ ::=⟨szöveg⟩

⟨Alcímek⟩ ::=⟨Alcím⟩⟨Alcímek⟩|⟨Alcím⟩

⟨Alcím⟩ ::= : ⟨Alcím adata⟩|λ

⟨Alcím adata⟩ ::=⟨szöveg⟩

⟨Párhuzamos címek⟩ ::=⟨Párhuzamos cím⟩⟨Párhuzamos címek⟩|⟨Párhuzamos cím⟩

⟨Párhuzamos cím⟩ ::=⟨Párhuzamos cím törzse⟩⟨Alcímek⟩⟨Szerzőségi közlés⟩

⟨Párhuzamos cím törzse⟩ ::= = ⟨Párhuzamos cím adata⟩|λ

⟨Párhuzamos cím adata⟩ ::=⟨szöveg⟩

⟨Egyéb címadatok⟩ ::=⟨Egyéb címadat⟩⟨Egyéb címadatok⟩|⟨Egyéb címadat⟩

⟨Egyéb címadat⟩ ::= : ⟨Egyéb címadat adata⟩|λ

⟨Egyéb címadat adata⟩ ::=⟨szöveg⟩

⟨Szerzőségi közlés⟩ ::= / ⟨Szerzőségi közlés adata⟩|λ

⟨Szerzőségi közlés adata⟩ ::=⟨Szerzők csoportjai⟩⟨További közreműködők csoportjai⟩

⟨Szerzők csoportjai⟩ ::=⟨Szerzők csoportja⟩ ; ⟨Szerzők csoportjai⟩|
⟨Szerzők csoportja⟩

⟨Szerzők csoportja⟩ ::=⟨Szerző⟩, ⟨Szerzők csoportja⟩|⟨Szerző⟩

⟨Szerző⟩ ::=⟨szöveg⟩

⟨További közreműködők csoportjai⟩ ::=⟨További közreműködők csoportja⟩ ;
⟨További közreműködők csoportjai⟩|
⟨További közreműködők csoportja⟩

⟨További közreműködők csoportja⟩ ::=⟨További közreműködő⟩,
⟨További közreműködők csoportja⟩|
⟨További közreműködő⟩

⟨További közreműködő⟩ ::=⟨szöveg⟩

⟨Kiadás⟩ ::=. – ⟨Kiadás adata⟩|λ

⟨Kiadás adata⟩ ::=⟨Kiadásmegjelölés⟩⟨Kiadás módját megadó szöveg⟩

⟨Kiadásmegjelölés⟩ ::=⟨Kiadásszám⟩. ⟨'Kiadás' vagy 'Kiad.'⟩|λ

⟨Kiadásszám⟩ ::=⟨arab szám⟩

⟨'Kiadás' vagy 'Kiad.'⟩ ::=Kiadás|Kiad.

⟨Kiadás módját megadó szöveg⟩ ::=⟨szöveg⟩

⟨Megjelenés⟩ ::=. – ⟨Megjelenés adata⟩

⟨Megjelenés adata⟩ ::=⟨Hely—Kiadó–csoportok⟩, ⟨Megjelenés ideje⟩

⟨Hely—Kiadó–csoportok⟩ ::=⟨Hely—Kiadó–csoport⟩ ; ⟨Hely—Kiadó–csoportok⟩|
⟨Hely—Kiadó–csoport⟩

⟨Hely—Kiadó–csoport⟩ ::=⟨Megjelenési helyek⟩ : ⟨Kiadók⟩

⟨Megjelenési helyek⟩ ::=⟨Megjelenési hely⟩ ; ⟨Megjelenési helyek⟩|⟨Megjelenési hely⟩

⟨Megjelenési hely⟩ ::=⟨szöveg⟩

⟨Kiadók⟩ ::=⟨Kiadó⟩ : ⟨Kiadók⟩|⟨Kiadó⟩

⟨Kiadó⟩ ::=⟨szöveg⟩

⟨Megjelenés ideje⟩ ::=⟨Megjelenési idő adata⟩⟨Copyright⟩

⟨Megjelenési idő adata⟩ ::=⟨Nyitott évszám⟩|⟨Zárt évszám⟩

⟨Nyitott évszám⟩ ::=⟨Zárt évszám⟩–

⟨Zárt évszám⟩ ::=⟨évszám⟩|⟨Post⟩|⟨Ante⟩|⟨Circa⟩|⟨Kérdőjeles évszám⟩|
⟨Sine anno⟩

⟨Post⟩ ::=[post ⟨évszám⟩]

⟨Ante⟩ ::=[ante ⟨évszám⟩]

⟨Circa⟩ ::=[cca ⟨évszám⟩]|[ca ⟨évszám⟩]

⟨Kérdőjeles évszám⟩ ::=⟨számjegy⟩⟨számjegy⟩⟨számjegy⟩?

⟨Copyright⟩ ::=, cop.⟨évszám⟩|λ

⟨Sine anno⟩ ::=[s.a.]

⟨Terjedelem⟩ ::=. – ⟨Terjedelem adata⟩|λ

⟨Terjedelem adata⟩ ::=⟨Oldalszám adat⟩⟨Tábla⟩⟨Illusztráció⟩⟨Méret⟩⟨Melléklet⟩

⟨Oldalszám adat⟩ ::=⟨Egyszerű oldalszám⟩|⟨Kétszer induló oldalszám⟩|
⟨Összegzett oldalszám⟩|⟨Főrészes oldalszám⟩|
⟨Futó oldalszám⟩|⟨Bizonytalan oldalszám⟩

⟨Egyszerű oldalszám⟩ ::=⟨arab szám⟩ ⟨Jelzet⟩

⟨Jelzet⟩ ::=⟨Pagina⟩|⟨Folio⟩|⟨Columna⟩

⟨Pagina⟩ ::=p.

⟨Folio⟩ ::=fol.

⟨Columna⟩ ::=col.

⟨Kétszer induló oldalszám⟩ ::=⟨római szám⟩, ⟨Egyszerű oldalszám⟩

⟨Összegzett oldalszám⟩ ::=[⟨Egyszerű oldalszám⟩]

⟨Főrészes oldalszám⟩ ::=⟨Egyszerű oldalszám⟩, ⟨Összegzett oldalszám⟩

⟨Futó oldalszám⟩ ::=⟨arab szám⟩–⟨Egyszerű oldalszám⟩

⟨Bizonytalan oldalszám⟩ ::=kb. [⟨Egyszerű oldalszám⟩]

⟨Tábla⟩ ::=, ⟨Tábla adata⟩|λ

⟨Tábla adata⟩ ::=⟨Táblaszám–megjelölés⟩ ⟨Táblajelzet⟩

⟨Táblaszám–megjelölés⟩ ::=⟨arab szám⟩|[⟨arab szám⟩]

⟨Táblajelzet⟩ ::=⟨Táblalap⟩|⟨Táblaoldal⟩

⟨Táblalap⟩ ::=t.fol.

⟨Táblaoldal⟩ ::=t.

⟨Illusztráció⟩ ::= : ⟨Illusztráció adata⟩|λ

⟨Illusztráció adata⟩ ::=⟨Illusztráció–jel⟩⟨Kiegészítő szöveg⟩

⟨Illusztráció–jel⟩ ::=ill.

⟨Kiegészítő szöveg⟩ ::=, ⟨szöveg⟩|λ

⟨Méret⟩ ::= ; ⟨Méret adata⟩|λ

⟨Méret adata⟩ ::=⟨Egydimenziós méret⟩|⟨Kétdimenziós méret⟩

⟨Egydimenziós méret⟩ ::=⟨arab szám⟩⟨Mértékegység⟩

⟨Mértékegység⟩ ::=cm|mm

⟨Kétdimenziós méret⟩ ::=⟨arab szám⟩ × ⟨Egydimenziós méret⟩

⟨Melléklet⟩ ::= + ⟨Melléklet adata⟩|λ

⟨Melléklet adata⟩ ::=⟨Mellékletjel⟩⟨Kiegészítő szöveg⟩⟨Mellékletterjedelem⟩

⟨Mellékletjel⟩ ::=mell.

⟨Mellékletterjedelem⟩ ::= (⟨Terjedelem adata⟩)|λ

⟨Sorozat⟩ ::=. – ⟨Sorozat adata⟩|λ

⟨Sorozat adata⟩ ::=⟨Sorozati szerkezet⟩. ⟨Sorozat adata⟩|⟨Sorozati szerkezet⟩

⟨Sorozati szerkezet⟩ ::=(⟨Fősorozat⟩)

⟨Fősorozat⟩ ::=⟨Alsorozat⟩. ⟨Fősorozat⟩|⟨Alsorozat⟩

⟨Alsorozat⟩ ::=⟨Címek és szerzőségi közlések⟩⟨ISSN⟩⟨Sorozati szám⟩

⟨ISSN⟩ ::=, ISSN ⟨ISSN adata⟩|λ

⟨ISSN adata⟩ ::=⟨Négyjegyű rész⟩ – ⟨Négyjegyű rész⟩

⟨Négyjegyű rész⟩ ::=⟨számjegy⟩⟨számjegy⟩⟨számjegy⟩⟨számjegy⟩

⟨Sorozati szám⟩ ::= ; ⟨arab szám⟩.|λ

⟨Megjegyzés⟩ ::=. – ⟨Megjegyzés adata⟩|λ

⟨Megjegyzés adata⟩ ::=⟨szöveg⟩

⟨Terjesztés⟩ ::=. – ⟨Terjesztés adata⟩|λ

⟨Terjesztés adata⟩ ::=⟨ISBN⟩⟨Kivitel⟩⟨Ár⟩

⟨ISBN⟩ ::=ISBN ⟨ISBN adata⟩ |λ

⟨ISBN adata⟩ ::=⟨pontosan tízjegyű arab szám⟩

⟨Kivitel⟩ ::=kötve|fűzve|kötött|fűzött|λ

⟨Ár⟩ ::= : ⟨Ár adata⟩|λ

⟨Ár adata⟩ ::=⟨arab szám⟩, – Ft|ár nélkül

2. melléket
Könyvek bibliográfiai leírása – a levezetési lánc

⟨Könyvek bibliográfiai leírása⟩ ::=
⟨Címek és szerzőségi közlések⟩⟨Kiadás⟩⟨Megjelenés⟩⟨Terjedelem⟩
⟨Sorozat⟩⟨Megjegyzés⟩⟨Terjesztés⟩ ::=
⟨Címek és szerzőségi közlések⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Sorozat⟩⟨Megjegyzés⟩⟨Terjesztés⟩ ::=
⟨Címek és szerzőségi közlések⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Megjegyzés⟩⟨Terjesztés⟩ ::=
⟨Címek és szerzőségi közlések⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Cím és szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Cím⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Párhuzamos címek⟩⟨Alcímek⟩⟨Egyéb címadatok⟩
⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Párhuzamos cím⟩⟨Alcímek⟩⟨Egyéb címadatok⟩
⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Párhuzamos cím törzse⟩⟨Alcímek⟩⟨Szerzőségi közlés⟩⟨Alcímek⟩
⟨Egyéb címadatok⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Alcímek⟩⟨Szerzőségi közlés⟩⟨Alcímek⟩⟨Egyéb címadatok⟩
⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Alcím⟩⟨Szerzőségi közlés⟩⟨Alcímek⟩⟨Egyéb címadatok⟩
⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Szerzőségi közlés⟩⟨Alcímek⟩⟨Egyéb címadatok⟩
⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Alcímek⟩⟨Egyéb címadatok⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Alcím⟩⟨Egyéb címadatok⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Egyéb címadatok⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Egyéb címadat⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcímek⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=

⟨Főcím⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩⟨Alcímek⟩⟨Szerzőségi közlés⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩⟨Alcímek⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩⟨Alcím⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩⟨Megjelenés⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenés adata⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Hely—Kiadó-csoportok⟩, ⟨Megjelenés ideje⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Hely—Kiadó-csoport⟩, ⟨Megjelenés ideje⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési helyek⟩ : ⟨Kiadók⟩, ⟨Megjelenés ideje⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadók⟩, ⟨Megjelenés ideje⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenés ideje⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩⟨Copyright⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩⟨Terjedelem⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩. – ⟨Terjedelem adata⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩. – ⟨Oldalszám adat⟩⟨Tábla⟩⟨Illusztráció⟩⟨Méret⟩⟨Melléklet⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩. – ⟨Oldalszám adat⟩⟨Illusztráció⟩⟨Méret⟩⟨Melléklet⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩. – ⟨Oldalszám adat⟩⟨Illusztráció⟩⟨Méret⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩,
⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩⟨Illusztráció⟩⟨Méret⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció adata⟩⟨Méret⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩⟨Kiegészítő szöveg⟩⟨Méret⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩ : ⟨Méret adata⟩⟨Terjesztés⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩ : ⟨Egydimenziós méret⟩. – ⟨Terjesztés adata⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩ : ⟨Egydimenziós méret⟩. – ⟨ISBN⟩⟨Kivitel⟩⟨Ár⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩ : ⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩⟨Ár⟩ ::=
⟨Főcím adata⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ :
⟨Illusztráció–jel⟩ : ⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨Megjelenési hely⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ : ⟨Illusztráció–jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨Kiadó⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ : ⟨Illusztráció–jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨Megjelenési idő adata⟩. – ⟨Egyszerű oldalszám⟩ : ⟨Illusztráció–jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨Egyszerű oldalszám⟩ : ⟨Illusztráció–jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ ⟨Jelzet⟩ : ⟨Illusztráció–jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=

⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ ⟨Pagina⟩ : ⟨Illusztráció-jel⟩ :
⟨Egydimenziós méret⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ p. : ⟨Illusztráció-jel⟩ :
⟨arab szám⟩⟨Mértékegység⟩. – ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm. –
ISBN ⟨ISBN adata⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm. –
ISBN ⟨pontosan tízjegyű arab szám⟩ ⟨Kivitel⟩ : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm. –
ISBN ⟨pontosan tízjegyű arab szám⟩ fűzött : ⟨Ár adata⟩ ::=
⟨szöveg⟩. – ⟨szöveg⟩ : ⟨szöveg⟩, ⟨évszám⟩. – ⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm. –
ISBN ⟨pontosan tízjegyű arab szám⟩ fűzött : ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében . – ⟨szöveg⟩ : ⟨szöveg⟩,
⟨évszám⟩.–⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm.–ISBN ⟨pontosan tízjegyű arab szám⟩ fűzött :
ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében .– Bátaszék : ⟨szöveg⟩,
⟨évszám⟩.–⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm.–ISBN ⟨pontosan tízjegyű arab szám⟩ fűzött :
ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében.– Bátaszék : II. Géza
Gimnázium , ⟨évszám⟩. – ⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm. – ISBN
⟨pontosan tízjegyű arab szám⟩ fűzött : ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében.– Bátaszék : II. Géza
Gimnázium , 1993 .–⟨arab szám⟩ p. : ill. : ⟨arab szám⟩cm.–ISBN ⟨pontosan tízjegyű arab szám⟩
fűzött : ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében.– Bátaszék : II. Géza
Gimnázium , 1993 .– 192 p. : ill. : ⟨arab szám⟩cm. – ISBN ⟨pontosan tízjegyű arab szám⟩
fűzött : ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében.– Bátaszék : II. Géza
Gimnázium , 1993 .– 192 p. : ill. : 20 cm . - ISBN ⟨pontosan tízjegyű arab szám⟩ fűzött :
ár nélkül ::=
A bátaszéki II. Géza Gimnázium jubileumi évkönyve az 1992-93. iskolai évről
az iskola fennállásának 30., Bátaszék alapításának 851. évében.– Bátaszék : II. Géza
Gimnázium , 1993 .– 192 p. : ill. : 20 cm . - ISBN 9630436477 fűzött : ár nélkül

3. melléket
ETO – egy lehetséges grammatika

⟨ETO-szám⟩ ::=⟨ETO-szám⟩+⟨ETO-szám⟩|
⟨ETO-szám⟩/⟨ETO-szám⟩|
⟨ETO-szám⟩/.⟨ETO-szám⟩|
⟨ETO-szám⟩:⟨ETO-szám⟩|
⟨ETO-szám⟩::⟨ETO-szám⟩|

|⟨ETO-szám⟩||
⟨ETO-szám⟩*⟨Idegen jelzet⟩|
⟨Főtáblázati szám⟩=⟨Nyelvi alosztás⟩|
⟨Főtáblázati szám⟩(=⟨Népi alosztás⟩)|
⟨Főtáblázati szám⟩(0⟨Formai alosztás⟩)|
⟨Főtáblázati szám⟩(⟨Földrajzi alosztás⟩)|
⟨Főtáblázati szám⟩"⟨Időalosztás⟩"|
⟨Főtáblázati szám⟩−03⟨Anyagalosztás⟩|
⟨Főtáblázati szám⟩−05⟨Személycsoport-alosztás⟩|
⟨Főtáblázati szám⟩.00⟨Szempontalosztás⟩|
⟨Főtáblázati szám⟩−⟨Korlátozottan közös alosztás (1)⟩|
⟨Főtáblázati szám⟩.0⟨Korlátozottan közös alosztás (2)⟩|
⟨Főtáblázati szám⟩'⟨Korlátozottan közös alosztás (1)⟩|
⟨Főtáblázati szám⟩⟨Alfabetikus alosztás⟩|⟨Főtáblázati szám⟩

⟨Főtáblázati szám⟩ ::=⟨Számjegy⟩|⟨Számjegy⟩⟨Számjegy⟩|⟨Számjegy⟩⟨Számjegy⟩⟨Számjegy⟩|
⟨Számjegy⟩⟨Számjegy⟩⟨Számjegy⟩.⟨Főtáblázati szám⟩

⟨Számjegy⟩ ::=0|1|2|3|4|5|6|7|8|9

⟨Nyelvi alosztás⟩ ::=⟨Főtáblázati szám⟩

⟨Népi alosztás⟩ ::=⟨Főtáblázati szám⟩

⟨Formai alosztás⟩ ::=⟨Pozitív számjegy⟩|⟨Pozitív számjegy⟩⟨Számjegy⟩|
⟨Pozitív számjegy⟩⟨Számjegy⟩.⟨Főtáblázati szám⟩|
:⟨Főtáblázati szám⟩

⟨Pozitív számjegy⟩ ::=1|2|3|4|5|6|7|8|9

⟨Földrajzi alosztás⟩ ::=⟨Főtáblázati szám⟩

⟨Időalosztás⟩ ::=⟨Kronológiai idő⟩/⟨Kronológiai idő⟩|
.../⟨Kronológiai idő⟩|
⟨Kronológiai idő⟩/...|
⟨3−7⟩⟨Fenomenológiai idő⟩

⟨Kronológiai idő⟩ ::=⟨Előjel⟩⟨Időadat⟩

⟨Előjel⟩ ::= + | − |λ

⟨Időadat⟩ ::=⟨Év⟩⟨Hó⟩⟨Nap⟩

⟨Év⟩ ::=⟨0−2⟩|⟨0−2⟩⟨Számjegy⟩|⟨0−2⟩⟨Számjegy⟩⟨Számjegy⟩|
⟨0−2⟩⟨Számjegy⟩⟨Számjegy⟩⟨Számjegy⟩

⟨0−2⟩ ::=0|1|2

⟨Hó⟩ ::=.01|.02|.03|.04|.05|.06|.07|.08|.09|.10|.11|.12|λ

⟨Nap⟩ ::=.01|.02|.03|.04|.05|.06|.07|.08|.09|.10|.11|.12|.13|.14|.15|.16|
.17|.18|.19|.20|.21|.22|.23|.24|.25|.26|.27|.28|.29|.30|.31|λ

⟨3−7⟩ ::=3|4|5|6|7

⟨Anyagalosztás⟩ ::=⟨Számjegy⟩|⟨Számjegy⟩.⟨Főtáblázati szám⟩

⟨Személycsoport-alosztás⟩ ::=⟨Számjegy⟩|⟨Számjegy⟩.⟨Főtáblázati szám⟩

⟨Szempontalosztás⟩ ::=⟨Számjegy⟩|⟨Számjegy⟩.⟨Főtáblázati szám⟩

⟨Korlátozottan közös alosztás (1)⟩ ::=⟨Főtáblázati szám⟩

⟨Korlátozottan közös alosztás (2)⟩ ::=⟨Pozitív számjegy⟩|⟨Pozitív számjegy⟩⟨Számjegy⟩|
⟨Pozitív számjegy⟩⟨Számjegy⟩.⟨Főtáblázati szám⟩

4. melléket
A VIVALDI leírása

A program hívásakor négy (illetve az úgynevezett *debug* üzemmódban öt) file-nevet adunk meg. Mind a négy (öt) file közönséges ASCII szövegfile. Jellemzőik:

- **bnf-file** alább részletesen ismertetendő file, mely a BNF definíciókat tartalmazza
- **elemzendő input** az input-file soronként egy elemzendő egységet tartalmaz (elemzendő egység az, amelyről a bnf-file-ban található BNF szól)
- **vivaldi output** a *VIVALDI* program output listája
- **hibafile** a sikertelen elemzéseket tartalmazó file
- **debug-file** olyan outputfile, amely az elemzés nyomonkövetését segíti

A *VIVALDI* 2.2-es verziójában az opciók az alábbiak:

A Angol nyelvű built-in-készlet

A *VIVALDI*-nak beépített (built-in) szimbólumkészlete van, amely a BNF tovább nem elemzendő (atomi) egységeit — kvázi terminálisait — jeleníti meg. A készlet megvan magyar és angol változatban. A *VIVALDI* alapértelmezés szerint a magyart használja, nemzetközi felhasználás illetve ékezetes karakterekkel kapcsolatos problémák esetén azonban szükség lehet az angol készletre.

C⟨készlet⟩ a .Betű halmaz megválasztása ⟨készlet⟩ megengedett értékei:
CWI⟨iiiooo⟩, 852DOS, 852WINDOWS, MaTeX, Angol — ahol ⟨iiiooo⟩
a nagy hosszú Í illetve Ő ASCII kódja [default: CWI141139]

A .Betű halmaz a *VIVALDI* egyik tovább nem elemzendő nemterminálisa (v. ö. a ⟨szöveg⟩-ről mondottakkal), ennek többféle beállítása lehet célszerű. A legegyszerűbb az *Angol* halmaz, ennek választásakor csak az angol ábécé betűi számítanak betűknek. A valamelyik másik opció választásával a széles körben elterjedt ékezetes betűkészletekből választhatunk. Megjegyezzük, hogy a felsoroltakkal a megalkotható készletek nem merültek ki, a felhasználó a BNF-ben természetesen elkészítheti saját *Betű* halmazát.

DB⟨string⟩ egységhatároló string: az elemzendő csomag kezdetét jelöli

DE⟨string⟩ egységhatároló string: az elemzendő csomag végét jelöli

A fenti két opcióra akkor van szükség, amikor nem a teljes inputfile-t kell elemezni, hanem annak csak egyes részleteit. A *VIVALDI* az elemzendő részleteket *csomag*oknak nevezi. A határoló stringek speciális, az inputban másutt elő nem forduló jelsorozatok. Egy-egy csomagot a fenti két határolóstringgel kell közrefogni.

E[⟨sorvégjel⟩]empty-line-mód: üres sorok választják el az egységeket ⟨sorvégjel⟩ a Carriage Return + Line Feed helyére értendő

Amikor a teljes inputfile elemzésre kerül, akkor is tudnunk kell, hogy melyek az abban található elemzendő egységek. A /E opció arra utal, hogy az inputfile-ban az egységeket üres sorok választják el. A sorvégek ilyenkor szükség esetén behelyettesítődnek. A belyehettesített jelsorozat ugyanúgy elemzés alá kerül, mint az elemzendő input többi részlete.

H⟨rejtettjel⟩ a BNF-fej előtt elhelyezhető egy karakter, mely a fej rejtettségét jelöli [default: *]

A *VIVALDI* voltaképpeni munkája az, hogy az elemzés során a sikeresen elemzett jelsorozatokhoz hozzákapcsolja azt a nemterminálist (más szóval BNF-fejet), amelyen keresztül a jelsorozat a Chomsky-féle generatív grammatikában levezetődött. Vannak azonban olyan esetek, amikor ez a hozzákapcsolás fölösleges, például azért, mert más hozzákapcsolásokkal együtt tautológiákra vezetne. Ilyenkor a BNF-fejnek rejtettnek kell maradnia. A rejtettség tényét a BNF írója egy a fej elé szúrt speciális karakter — például csillag — segítségével közli a *VIVALDI*-val.

L⟨sorvégjel⟩ a nyitó-/csukójelek közötti ⟨sorvégjel⟩-ből soremelés lesz [default nincs]

Segédeszköz az output sorokra tördelése céljából. Megfelelően elhelyezett speciális karakterekkel mód van arra, hogy soremeléseket helyezzünk el az outputban.

N NUL-mód: az egységhatárolókon kívüli szövegrész nem kerül át az outputba

Figyelemmel arra, amit a /DB, /DE opcióknál mondtunk két eset lehetséges: az inputnak azok a részei, amlyek a *csomag*okon kívül esnek, átkerülnek illetve nem kerülnek át az outputba. A /N opció ezt szabályozza.

P suppress-mód: nincs egyenlőségjeles tagolás az outputban

Az elemzendő egységekből származó output részleteket a *VIVALDI* egyenlőségjelekből álló jelzettel választja el egymástól. Ez sok szempontból — például a látvány szempontjából — előnyös. Elképzelhető azonban olyan további feldolgozás, amikor a következő feldolgozóprogramot az egyenlőségjelek zavarnák. A /P opcióval futó *VIVALDI* az egyenlőségjeleket nem helyezi el az outputban.

Q quiet mód: nincs képernyőre írás futás közben

S⟨szeparátor⟩ a BNF-fejben a ::= előtt elhelyezhető nyitó- és csukójelek szeparátora [default: =]

Egy sikeresen elemzett grammatikai egységet a *VIVALDI* — igény esetén — az elején és a végén címkéz fel. Ezek a címkék a nyitó- illetve csukójelek. E jeleket (melyek több karakterből is állhatnak) egy speciális karakter — például az egyenlőségjel — határolja.

T teszt-mód: megáll az első elutasított tételnél

A *VIVALDI* normál üzemmódja az, hogy hiba nélkül feldolgozza az input-file összes egységét. Addig azonban, amíg a BNF nincs készen, a *VIVALDI* a BNF íróját ezzel az opcióval támogatja.

V[⟨méret⟩] nyomkövető mód: a hibás inputegységből ⟨méret⟩ számú byte megy a hibalistába [default: mind átkerül]

A teljes hibalista olykor rendkívül hosszú lehet. A /V opcióval a felhasználó ezt csökkentheti.

X⟨extrák⟩ a .Betű halmaz kiegészítése ASCII kódokkal — pl.: /X045039 kötőjellel és aposztróffal egészít ki

Korábban láttuk, hogy a .Betű halmaz megválasztható, a /X opcióval további karakterek tehetők betűkké.

Z ugyanaz, mint az E opció szóközzel

A /E opció a szóközzel, mint sorvéghelyettesítővel olvashatósági okokból a /Z alakot ölti.

? angol nyelvű helpszöveg kiírása

Néhány példa a *VIVALDI* hívására: (A magyarázatok a példa *UTÁN* állnak.)

```
vivaldi szocikk.bnf szotar.txt szotar.lst szotar.hib
```

Ez a *VIVALDI* lehető legegyszerűbb hívása, a SZOTAR.TXT-ben az elemzendő egység a sor. Minden sor elemzésre kerül.

```
vivaldi szocikk.bnf szotar.txt -z szotar.lst szotar.hib
```

Fentiek úgy módosulnak, hogy a SZOTAR.TXT-ben az elemzendő egységeket üres sorok határolják. Az egységeken belüli sorvégek szóközre cserélődnek.

```
vivaldi szocikk.bnf szotar.txt -z szotar.lst szotar.hib -T szotar.deb
```

A *VIVALDI* nyomkövető listát készít és az első elutasított tételnél megáll.

```
vivaldi szocikk.bnf /cCWI140167 /X046033063 szotar.txt szotar.lst szotar.hib
```

A *VIVALDI* CWI kódot vár azzal a módosítással, hogy a nagy hosszú Í kódja 140, a nagy hosszú Ő-é pedig 167. A szokásos betűk mellett betűnek tekinti a pontot, a felkiáltójelet és a kérdőjelet is.

```
vivaldi szocikk.bnf -DB -DE -P szotar.txt szotar.lst szotar.hib
```

SZOTAR.TXT-nek csak a kapcsos zárójelbe tett részletei kerülnek elemzésre. A {...} jeleken kívüli részletek változtatás nélkül átkerülnek SZOTAR.LST-be. SZOTAR.LST-ben nem lesznek egyenlőségjeles tagolók.

```
vivaldi @szocikk.vez /V50
```

A *VIVALDI* a parancssor-paramétereket a SZOCIKK.VEZ file-ból veszi, az ottaniak mellé a /V50 opció társul, ez utóbbi jelentése szerint hibás egységből annak első 50 byte-ja kerül majd a hibafile-ba. Tételezzük fel, hogy SZOCIKK.VEZ tartalma az alábbi:

```
//   Hungarian Entry Analyzer Package   (c) MorphoLogic
//   Author:         Charles the Modest
//   Launch date:    June, 29. 1997
szocikk.bnf          //   BNF grammar
//   NOTE!!  Angle brackets are not to be given
//             via command line!
/DB<unit>            //   SGML opener
/DE</unit>           //   SGML closer
/P                   //   Without === separators
/N                   //   Parts out of units do not come
szotar.txt           //   Input text
szotar.lst           //   Output listing
szotar.hib           //   Error listing
//   End of File SZOCIKK.VEZ
```

Ekkor a SZOCIKK.VEZ file-ból a nemüres és nemkomment sorok a parancssorba lépnek, a

```
vivaldi @szocikk.vez /V50
```

hívás (majdnem[29]) egyenértékű lesz az alábbival:

```
vivaldi szocikk.bnf /DB<unit> /DE</unit> /P /N szotar.txt szotar.lst
                                                    szotar.hib /V50
```

Utolsó példánk egy hibás hívás:

```
vivaldi szocikk.bnf /DB(unit start) /DE(unit stop) szotar.txt szotar.lst
                                                              szotar.hib
```

Ebben a parancssorban a paraméterek száma 8, ebből csak kettő opció, ezek: /DB(unit és /DE(unit. A feltételezett szándék szerinti helyes parancssor az alábbi lett volna:

```
vivaldi szocikk.bnf /DB(unit\ start) /DE(unit\ stop) szotar.txt szotar.lst
                                                              szotar.hib
```

A BNF-file-ban # (hashmark) karakterek segítségével komment sorokat lehet elhelyezni. A nemkomment sorok tartalmazzák a BNF nyelvtant az alábbiak szerint.

Az egyes BNF-definíciókat üres sorok választják el egymástól. (Több egymásutáni üres sor egy üres sornak számít.) Egy BNF-definíció alakja az alábbi:

$[rejtettjel]\langle fej\rangle[=nyitó\ címke[=csukó\ címke]]$ ::=

$\langle tényező\rangle[\langle tényező\rangle...\langle tényező\rangle]$

$[|\langle tényező\rangle[\langle tényező\rangle...\langle tényező\rangle]|\ ...\ |\langle tényező\rangle[\langle tényező\rangle...\langle tényező\rangle]]$

A dőntött betűk a behelyettesítendőket, a [ ] | karakterek és a pontok a metaszintaxis elemei, míg a többi betű a BNF-definíció konstans része.

Felhívjuk a figyelmet néhány körülményre:

1. Az így megadott Chomsky-értelemben környezetfüggetlen nyelvtannak λ-mentesnek kell lennie[30].

2. Bevezetünk egy fogalmat: Egy tetszőleges T ábécé fölötti T^* térben legyenek p és q T^*-beli elemek. Ekkor azt mondjuk, hogy p a q-ból egy lépésben elhagyással nyerhető, ha létezik q-nak olyan $q = q_1q_2q_3$ felosztása, hogy $p = q_1q_3$. A q bármely darabja lehet λ.

   Azt mondjuk továbbá, hogy p a q-ból elhagyással nyerhető, ha létezik T^*-beli elemeknek egy olyan $r_1, r_2, ..., r_n$ sorozata, hogy $r_1 = p$ és $r_n = q$ és minden előző r a következőből egy lépésben elhagyással nyerhető.

   Könnyen belátható, hogy az elhagyással nyerhető reláció parciális rendezés. Ennek a fogalomnak a felhasználásával azt a követelményt támasztjuk az egyes BNF-definíciókkal szemben, hogy a *vertical bar* karakterekkel elválasztott diszjunktív tagok olyan sorrendben álljanak, hogy t_1 tag ne előzze meg t_2 tagot, ha t_1 tag t_2 tagból elhagyással nyerhető. Mivel a reláció parciális rendezés, ez a követelmény mindig teljesíthető.

   (A fentiek ellenére előfordulhat azonban, hogy az elemző úgynevezett *garden path* szituációba[31] keveredik: sikeres elemzésből nem tud visszalépni. Ennek a kivédésére egyelőre csak a BNF alkalmas átírása szolgál — már persze, hogyha ez elméletileg lehetséges[32].)

3. A jellegzetes

$$\langle x\rangle ::= \langle y\rangle\langle x\rangle|\langle y\rangle$$

   alakú rekurzív definíciók kimeríthetik a *VIVALDI* stack-mélységét.
   A *VIVALDI* egy $\langle x\rangle ::= \langle y\rangle\langle x\rangle|\langle y\rangle$ alakú fejdefiníció esetén azokat az $\langle x\rangle$-eket képes elemzni, amelyek kevesebb mint 80 $\langle y\rangle$-ból állnak. Amennyiben az inputban ennél hosszabb egységek várhatók, javasoljuk a grammatika olyan átírását, mely az $\langle x\rangle$-nek az $\langle y\rangle$-nál nagyobb egységekből álló közbülső struktúráját veszi figyelembe. Így például egy

⟨Szöveg⟩ ::= ⟨Jel⟩⟨Szöveg⟩|⟨Jel⟩

⟨Jel⟩ ::= ⟨.Betű⟩|⟨Szóköz⟩

⟨Szóköz⟩ ::= ⟨.Konstans ⟩

BNF-részlet helyett az alábbi átírást javasolhatjuk:

⟨Szöveg⟩ ::= ⟨Szó⟩⟨Szóköz⟩⟨Szöveg⟩|⟨Szó⟩

⟨Szó⟩ ::= ⟨.Betű⟩⟨Szó⟩|⟨.Betű⟩

⟨Szóköz⟩ ::= ⟨.Konstans ⟩

Funkcióikat tekintve a BNF részleteiről az alábbiakat jelenthetjük ki:

- Az első BNF-definíció fej eleme a mondatszimbólum.
- A tényezők lehetnek beépített vagy felhasználó által definiált tényezők. A beépített tényezők elnevezése pont karakterrel kezdődik, felhasználó által definiált tényező neve nem kezdődhet ponttal.
- Felhasználó által definiált tényezőnek valahol a BNF-ben fejként is szerepelnie kell. A megfeleltetés a csúcsos zárójelek közötti névalak kis-/nagybetűhív illesztésével történik.
- A beépített nevek a következők (magyar illetve angol built-in-készlet (lásd A opció) mellett:)

.Nyitó	.Opener
.Csukó	.Closer
.Betű	.Letter
.Számjegy	.Cipher
.Római szám	.Roman numeral
.Arab szám	.Arabic numeral
.Konstans	.Constant
.Valamely	.Either
.Eddig	.Hitherto

Fentiek közül a .Konstans, a .Valamely és az .Eddig beépített név paraméterezhető, míg a többi nem paraméterezhető beépített név. Ez azt jelenti, hogy például a .Betű beépített nevet csak önmagában lehet használni, (v.ö. azonban C és X opciók!) míg például a .Konstans beépített név a csúcsos zárójelek között egyéb elemet is tartalmaz.

- A beépített nevek funkciója:

.Nyitó	A csúcsos nyitózárójelet jelenti, tulajdonképpen a .Konstans beépített név speciális esete.
.Csukó	A csúcsos csukózárójelet jelenti, tulajdonképpen a .Konstans beépített név speciális esete.
.Betű	Olyan nemterminális, amelyből egyetlen betű vezethető le. Az angol ábécé betűin kívüli további betűk a C és az X opciókkal adhatók meg.
.Számjegy	Olyan nemterminális, amelyből egyetlen számjegy vezethető le. A számjegyek a következők: 1 2 3 4 5 6 7 8 9 0.

.Római	Olyan nemterminális, amelyből római szám szám vezethető le.
.Arab szám	Olyan nemterminális, amelyből előjel nélküli egész arab szám vezethető le.
.Konstans	Konstans jelsorozat levezetésére alkalmas nemterminális. A konstans megadása a következő: A beépített név Konstans szava utáni egy szóköz és a csúcsos csukó zárójel közötti jelsorozat a szóbanforgó konstans. Maguk a csúcsos zárójelek nem adhatók meg .Konstans segítségével.
.Valamely	Alakja ⟨.Valamely jelsorozat⟩ A jelsorozatot egy szóköz választja el a Valamely szótól. A .Valamely sikeresen elemződik, ha a soronkövetkező jel az adott jelsorozatban megtalálható. Sikeres elemzés esetén mindig pontosan egy pozíciót halad.
.Eddig	Alakja ⟨.Eddig jelsorozat[jelsorozat|...|jelsorozat]⟩ Ennek a beépített névnek a hatására az elemzés addig halad az inputláncban, amíg az .Eddigben felsorolt jelsorozatok valamelyikével nem találkozik. (Hogyha egyik sincs az inputláncban, akkor az elemzés sikertelen.) Ha valamelyik jelsorozat vertical bar karaktert tartalmaz, akkor a jelsorozatot csúcsos zárójelbe kell tenni. Ilyenkor a csúcsos zárójelek nem számítanak bele a jelsorozatba.

- A *VIVALDI* 2.2 az inputot többféleképpen osztja egységekre:

  egy sort tekint egy egységnek (ez a default viselkedés)

  egy üres sortól üres sorig terjedő (lényegében egy bekezdésnyi) egységet definiál. (Lásd E illetve Z opció!) Az E opció alkalmazásával a sorvégeket a felhasználó tetszés szerinti nem szóköz karakterrel helyettesítheti (pl. a parancssorban elhelyezett /E+ opció minden sorvéget + jellel helyettesít), vagy akár elhagyhatja (üres /E opció). A Z opcióra a fent elmondottak érvényesek úgy, hogy a sorvégek helyére szóköz értendő.

  A D opció alkalmazásával egységhatárolók (delimiterek) döntik el, hogy az input mely részei kerülnek elemzés alá. A delimitereken kívüli inputrészletek vagy változtatás nélkül átkerülnek az outputba, vagy elvesznek (lásd N opció!).

Az elemzendő input minden így keletkezett egységéről elemzést készít az outputfile–ba. Egy output tétel (feltéve, hogy ezt a P opció felül nem bírálja) legalább egy

===:nnn:
alakú sort tartalmaz. Itt nnn helyén az inputfile–bani sorszám áll. Figyelemmel arra is, hogy a BNF–ben alkalmaztunk–e nyitó/csukó címkéket, az inputlánc azon részletei, melyek sikeres elemzés során egy nem rejtett fej értékei lettek, *nyitó címke érték csukó címke* alakban kikerülnek az outputfile–ba.

A *VIVALDI* 2.2 hibafile–t is készít arra az esetre, hogyha valamely elemzés sikertelen lenne. Ezek az outputfile–éhoz nagyon hasonló tételek az inputfile sorszáma mellett a hibafile–on belüli saját sorszámot is tartalmazzák, az output–információt pedig addig írják ki, ameddig a *VIVALDI* sikeres részelemzést tudott végrehajtani. Ennek segítségével valamint a *VIVALDI* ötödik filemegadása révén hibakeresési lehetőséget nyújt. Ez a file ugyanis (ha jelen van) egy kifejezett debug–outputot fog tartalmazni, melyből a felhasználó a bejárás minden ágát nyomonkövetheti.

Jegyzetek

[1] Nem kívánunk foglalkozni ezen a ponton a számítógépes implementáció kérdésével, azzal, hogy az absztrakt mezőt milyen valóságos számítástechnikai megoldás jeleníti meg.

[2] Egy apróságot azonban megemlítünk: Tekintsük a már idézett bibliográfiai leírás második és harmadik sorát az alábbi elhelyezkedésben:

```
évről az iskola fennállásának 30.,
Bátaszék alapításának 851. évében.
- Bátaszék : II. Géza Gimnázium, 1993.
- 192 p. : ill. ; 20 cm
```

Jóllehet a gondolatjel soreleji helyzete nem ütközik leírási szabványba, mégsem esztétikus. A probléma egy *kemény szóköznek* nevezett speciális jellel oldható meg. Ez a számítógépes karakterkészletnek egy minden más szempontból (megjelenés, szélesség, rendezési súly) a szóközzel egyenértékű karaktere, azzal a kivétellel, hogy kemény szóköznél nem lehet sort törni. (A számítógépek szokásos ASCII karakterkészletében ilyen karakter eleve nincs benne, a bibliográfiai leírást tördelő – feltehetőleg integrált könyvtári – programnak kell erre a célra egyet kiválasztania vagy felhasználnia. Az egyik lehetséges választás például az ASCII 255 karakter lehet.) A pont–gondolatjel kombinációban tehát kemény szóközt kell alkalmazni, és ekkor a gondolatjel sohasem kerülhet a sor elejére. A továbbiakban azonban az esetleges kemény szóközt nem különböztetjük meg a közönségestől.

[3] A λ-mentesítés a megfelelő fájlformátumban írt BNF nyelvtanokon automatikusan elvégezhető a szerző *Lambda* nevű programjával, amely letölthető a szerző honlapjáról: *http://isis.elte.hu/~csabay*

[4] „Sor"-on most egy ::=-höz tartozó egységet értünk.

[5] A formális nyelvek és automaták elméletében *célmondat*nak nevezik a tárgyalás során vizsgált jelsorozatot, mellyel kapcsolatban azt a kérdést kell eldönteni, hogy eleme-e a vizsgálat kezdetén rögzített nyelvnek. Esetünkben a célmondat a szóban forgó bibliográfiai leírás.

[6] A BNF jelen sorrendje miatt ténylegesen ez lesz az eredmény.

[7] Ennek hatására az egyenrangú darabok száma a példában ötről kettőre csökken. Annak felismerésére azonban, hogy a `II. Géza` egybefüggő szerkezet, egyszerű szintaktikai szempontot adni nem lehet.

[8] Ez azt jelenti, hogy ha egy könyvnek az a címe, hogy $E = m \cdot c^2$, akkor e címet a leírásba E[egyenlő]$m \cdot c^2$ alakban kell fölvenni.

[9] Elhelyezkedésében valahogy így: $Fc : fcac_1 : fcac_2 : fcac_3 = Pc : pcac_1 : pcac_2 : pcac_3$, ahol *Fc* a főcím, az *fcac*-k a főcím alcímei, *Pc* a párhuzamos cím, és a *pcac*-k a párhuzamos cím alcímei.

[10] Ha egy 200 oldalas könyvnek egy 30 oldalas melléklete van, és ez utóbbiban van egy négyfóliós újabb melléklet, akkor ezt a helyzetet a címleírás `200 p. + mell.(30 p. + mell.(4 fol.))` alakban fogja tükrözni.

[11] Hasonlítsuk össze ezt a formulát a közismert könyvtári adatbázis-kezelő rendszerek (pl. ALEPH, DIALOG) *AU=Smith, KW=logistics* avagy *Smith/AU, logistics/DE* formuláival!

[12] Szögletes zárójel közvetlenül, operátor használata nélkül nem illeszthető.

[13] Ismeretterjesztő jellegű, népszerű tárgyalásmódot jelentő formai alosztás.

[14] A szerző készített ilyen elemzőt, a példa valóságos, az elemző azzal utasította el a jelzetet, hogy a kötőjel utáni P betűvel nem tud mit kezdeni: ilyenfajta alosztás nincs.

[15] A VIVALDI 2.2 specifikációjában *Tihanyi László* és *Pál Miklós* (MorphoLogic) működtek közre. E közreműködésért a szerző ezúton mond köszönetet.

[16] A λ-mentesítést végző *Lambda* programról korábban már tettünk említést. Érdeklődők a programhoz mellékelt README fájlból tájékozódhatnak a program használatát illetően. Itt csak annyit jegyzünk meg, hogy a felhasználóknak a BNF-ben a lehetséges üres elemeket a diszjunkciók utolsó tagjaként kell elhelyezniük, és a művelet nem feltétlenül egy menetben zajlik le: a programot annyiszor kell futtatni egymás után, míg két egyforma outputot nem produkál.

[17] A fájl neve nem a fantázia szüleménye. Az eredeti BNF a *Lambda* programmal „kezelve" a negyedik „menetre" kerül a végleges állapotba.

[18] Figyeljünk fel néhány olyan apróságra, melyben az itt közölt elemzendő forma a korábbitól eltér: ilyen a bekezdés helyett pont-gondolatjel alkalmazása, az ISBN-szám folyamatos – szóközök nélküli – írásmódja. Ezekről a különbségekről korábban szóltunk.

[19] Ahhoz, hogy a VIVALDI-felhasználó idáig eljusson, a BNF aprólékos „hangolása" szükséges. Különös gondot kell fordítani például a szóközök előfordulására, a BNF-egységek jobb oldalán a diszjunktív tagok sorrendjének alkalmas megválasztására, és néhány olyan szempontra, amelyre itt nem tértünk ki, mert a részletezés akadály lett volna.

[20] Ha ezt nem tesszük, az egész output jelsorozat egyetlen sorba kerül.

[21] Felhasználói döntés kérdése, hogy a VIVALDI mit tekint érdektelennek. Ebben a példában a megjelenés éve érdektelen, más esetben nyilván szükség lehet erre az adatra.

[22] A MorphoLogic definiálta az úgynevezett *MoBiDic User Dictionary Format* szövegfájlszerkezetet, melynek leírását az érdeklődők a *MoBiDic kétnyelvű szótárak* CD-hez mellékelt füzetben (és természetesen a MoBiDic help szövegei között is) megtalálják.

[23] A # karakterek az SGML-zárójelek mellett majd soremelésre fognak cserélődni.

[24] Ez az output *nem* az iskolai szótár BNF-jével, hanem egy más, itt be nem mutatott BNF-fel jött létre.

[25] Legalább 32 MB-os, 1,3 MHz-es Pentium processzoros.

[26] Pl. C vagy Assembler.

[27] Az itt szereplő fogalmak definícióját lásd pl. *Varga László: Rendszerprogramok elmélete és gyakorlata.* – Budapest, Akadémiai, 1978. 342. p. és k.

[28] E-mail: *csabay@isis.elte.hu*

[29] Mint arra a példában elhelyezett komment is utal, az egyenértékűség azért csak „majdnem", mert a csúcsos zárójeleket parancssor útján a VIVALDI programnak (ahogy semmilyen más DOS programnak) átadni nem lehet.

[30] Emlékeztetjük az olvasót arra, hogy tetszőleges környezetfüggetlen nyelvtan λ-mentessé tehető; pl. a szerző *Lambda* programjával.

[31] Az elnevezés magyar alkalmazása *Prószéky Gábor*tól származik.

[32] A *garden path backtrack* bevezetését egy későbbi VIVALDI verzióban tervezzük.

Beérkezett: 1998. VII. 8-án.