

## **A hipertext számítógépes nyelvészeti alkalmazási lehetőségei Kutatási terv**

### 1. A kutatási előzmények

1984 óta több, mint félszáz publikációnk született a számítógépes nyelvészet, főként a konkordanciák és a hipertext számítógépes nyelvészeti alkalmazásának területén. A munka egyik fontos állomása a 2000-ben elkészített *Use of Hypertext in Information Science: Concepts, Systems, Models and Applications* című PhD disszertáció. Az újabb eredményeket *A hipertext számítógépes nyelvészeti alkalmazási lehetőségei* c. 2005-ös habilitációs dolgozat foglalja össze, amit azóta számos tanulmány és könyvrészlet követett. Az interneten 2008-ben tettük nyilvánosan hozzáférhetővé *Füst Milán* válogatott verseinek interaktív számítógépes konkordanciáját, 2009-ben pedig *Radnóti Mikós* verseinek interaktív számítógépes konkordanciáját.

### 2. A kutatási cél

Általánosan megfogalmazva kutatásunk tárgyát, az alapfeladat természetes nyelvű szövegek feldolgozása számítógépes támogatással, konkrétan pedig hipermediális, „webes” felületen hozzáférhető interaktív konkordancia-szótárak kifejlesztésével. Kutatásaink szempontjából a természetes nyelvű szövegek feldolgozásának alapproblémája úgy fogalmazható meg, hogy egy természetes nyelvű szöveg csak „a jéghegy csúcsa”, önmagában nem értelmezhető — ehhez további háttérismeretekre van szükség. A probléma általunk követett megoldása olyan komplex hipertext struktúra kialakítása az értelmezendő szövegek köré, amely a megértéshez és értelmezéshez szükséges legfontosabb információkat és kapcsolatokat tartalmazza interaktív konkordancia-szótárak formájában, amelyek mintegy „hidat” képeznek az értelmezendő szövegek és az értelmezéshez szükséges háttérismereteket modellező korpuszok között. Kutatásaink célja olyan webes felületen hozzáférhető hipertext (hipermédia) alapú „tudásbázisok” kialakítása és folyamatos bővítése (mind tartalmi szempontból, mind a tudásbázisban kialakított hipertext kapcsolatok, „linkek” számát illetően), amelyek a következő főbb komponensekből tevődnek össze:

- természetes nyelvű és multimedialis korpuszok,
- interaktív konkordancia-szótárak,
- a tudásbázis „belépési” pontjait képező alapszövegek,
- a tudásbázis interaktív használatát lehetővé tevő hipertextuális kapcsolatok,
- valamint interaktív keresési és statisztikai kiértékelő felületek.

### 3. A kutatás kivitelezésének terve

A kutatási célként kitűzött hipertext (hipermédia) alapú tudásbázisok kialakításának lépései nagy vonalakban a következők:

- a hipertextuális szövegértelmezéshez szükséges háttérismereteket modellező (digitális) korpuszok kialakítása és folyamatos bővítése; a hipertext alapú korpuszok struktúrájának megtervezése és webes megvalósítása;
- a korpuszok interaktív konkordancia-szótárának kialakítása;

- a vizsgálandó szövegek kiválasztása; a kiválasztott szövegek témájának megfelelő kulcsszórendszer kialakítása; a korpuszok és a konkordancia-szótárak folyamatos, célzott bővítése a kulcsszavakat hasonló jelentésben (és/vagy szöveggörnyezetben) tartalmazó természetes nyelvű és multimediális szövegekkel és ezekből származó konkordanciákkal;
- hipertext struktúra szervezése a kiválasztott szövegek köré a megfelelő kulcsszavak konkordanciáin keresztül;
- a különböző forrásokból származó szövegekből felépített komplex hiperszöveg szövegszerúségének (koherenciájának stb.) ellenőrzése, és szükség esetén értelmező megjegyzések, kommentárok beillesztése;
- kutatási hipotézisek felállítása és tesztelése; a kialakított tudásbázisok vizsgálata különböző szempontok szerint, például statisztikai mutatók definiálásával és vizsgálatával.

A fentiek alapján *A hipertext számítógépes nyelvészeti alkalmazási lehetőségei* kutatási projekt során célul kitűzött feladatok a következők:

vállalt feladat	egység	indikátor
elméleti háttér kutatása és eredmények közreadása	előadás tanulmány	4 4
digitális korpusz kialakítása	digitalizált szöveg (TXT)	
<i>Radnóti Miklós versei</i>		1
<i>Füst Milán válogatott versei</i>		1
<i>kiválasztott verseskötetek</i>		1-3
<i>kiválasztott prózai művek</i>		3-5
digitális korpusz struktúrájának kialakítása	metaadatrendszer (XML)	1
hipertext alapú korpusz kialakítása	hiperszöveg (HTML)	1
konkordancia szótárak létrehozása	számítógépes program (Java) és konkordancia szótár (HTML)	
<i>Radnóti Miklós verseinek konkordancia-szótára</i>		1
<i>Füst Milán válogatott verseinek konkordancia-szótára</i>		1
<i>kiválasztott művek konkordancia-szótára</i>		1-5
vizsgálandó szövegek kiválasztása és feldolgozása	kiválasztott szöveg	min. 4
<i>kulcsszórendszerek meghatározása</i>		
<i>új szövegek keresése és korpusz bővítése</i>		
<i>konkordancia-szótárak bővítése</i>		
<i>szövegek értelmezése</i>		
tudásbázisok interaktív keresése és (statisztikai) elemzése	számítógépes program (Java)	1

#### 4. A kutatás várható eredményei

A kutatási projekt fő eredményei a fentiekben részletezett hipertext (hipermédia) alapú tudásbázisok, amelyek magukban foglalják a következőket:

- metaadatrendszer és hipertext alapú korpusz;
- nyilvánosan, webes felületen hozzáférhető interaktív konkordancia szótárak (min. 3);
- Java alapú interaktív kereső és szövegstatisztikai elemző felület.

Emellett tervezzük a kutatási eredményeket összefoglaló monográfia kiadását is.