

*Kepa*

eman ta zabal zazu



euskal herriko  
unibertsitatea

universidad  
del país vasco

Informatika Fakultatea  
Facultad de Informática  
Apartado 649 / 649 Postakutxa  
Donostia - San Sebastián

FISS-I-47.1-LSI-91

HIZTEGI ELEBAKAR BATEKO DEFINIZIEN ANALISIA:  
PATROI-HIERARKIATAN OINARRITUTAKO ANALIZATZAI-  
LEAREN INPLEMENTAZIOA ETA LORTUTAKO EMAITZAK.

AGIRRE, E.  
ARREGI, X.  
ARTOLA, X.  
BASTARRIKA, J.R.  
EURARD, F.  
SARASOLA, K.

1991 URRIA

BARNE-TXOSTERA		IDENTIFIKAZIO-ORRIA	
GILTZA FISS-I-47.1-LSI-91		DATA 1991, URRIA	HIZKUNTZA EUSKARA
TXOSTEN-MOTA <input checked="" type="checkbox"/> IKERKUNTZA 1 <input type="checkbox"/> IRAKASKUNTZA 0 <input type="checkbox"/> TESIAK/TESIMAK 1 <input type="checkbox"/> BESTEAK 0		SAILA (K) LENGOAIA ETA SISTEMA INFORMATIKOAK ORRIALDE-ZENBAKIA 27      ERREFERENTZI-ZENBAKIA 6	
IZENBURUA HIZTEGI ELEBAKAR BATEKO DEFINIZIOEN ANALISIA: PATROI-HIERARKIATAN OINARRITUTAKO ANALIZATZAILEAREN INPLEMENTAZIOA ETA LORTUTAKO EMAITZAK			
EGILEA (K) E. Agirre, X. Arregi, X. Artola, J.R. Bastarrika, F. Eurard*, K. Sarasola			
EGILEARI DAGOKION ERAKUNDEA (DONOSTIAKO INFORMATIKA FAKULTATEA EZ IZATEKOTAN) * ENSEEIHT (TOULOUSE)			
OHOROKO ERAKUNDEAREN LAGUNTZAREKIN BURUTU DA LAMA (KONTRATUAREN ERREFERENTZIA)			
MEDAKUNTZA (BESTELA, MURRIZTALDIRIK GABE)			
BESTE OHARRAK			
LABURPENA <p>HIZTEGI ELEBAKAR BATEKO DEFINIZIOEN ANALISIA: PATROI-HIERARKIATAN OINARRITUTAKO ANALIZATZAILEAREN INPLEMENTAZIOA ETA LORTUTAKO EMAITZAK</p> <p>Txosten honetan hiztegi-definizioak sintaktikoki analizatzeko parser baten inplementazioa eta erabilera azaltzen dira.</p> <p>Analizatzailea patroi-hierarkiatan oinarritua dugu. Berau erabiliz analisi partzialak baina interpretazio semantiko zuzena emateko aukera eskaintzen dutenak egin daitezke.</p> <p>Analizatzailea inplementatu ondoren frantseseko hiztegi elebakar bateko definizioak analizatzeko erabili da. Txostenean lortutako emaitzen azalpena ere ematen da.</p>			
BERBA-GILTZ PROPOSATUAK JAKINTZAREN ESURAPENA HIZTEGIETATIK, ANALISI SINTAKTIKO-SEMANTIKOA, LENGOAIA NATURALAREN PROZESAMENDUA			
MEDAKUNTZA-ZERRENOA <input type="checkbox"/> OHOROKO ERANTSIKOKO MELBIDE OSKAR			

## AURKIBIDEA

1.- SARRERA.....	1
2.- PATROI-HIERARKIATAN OINARRITUTAKO ANALIZATZAILEAREN INPLEMENTAZIOA.....	2
2.1.- ZEHAZTAPENA .....	2
2.1.1.- Ingurunea.....	2
2.1.2.- Patroi-hierarkiaren adierazpidea.....	3
2.1.3.- Analisiaren emaitzak.....	6
2.1.4.- Katgoria logiko eta fisikoen arteko ezkonketa .....	7
2.2.- ALGORITMOA ETA INPLEMENTAZIOA .....	7
2.2.1.- Algoritmoa.....	7
2.2.2.- Modulu-diagrama eta paketeen erabilera .....	10
2.2.3.- Kanpo-funtzioak datu-base eta analizatzailearen arteko interface gisa .....	13
2.2.4.- Emaitzen tratamendua eta biltegitzea .....	14
2.3.- ERABILERA.....	17
2.3.1.- Batch moduan.....	17
2.3.2.- Modu elkarreragilean.....	18
2.3.3.- Nola gehitu katgoria bakunak.....	20
2.3.4.- Nola gehitu katgoria konposatuak.....	20
2.3.5.- Nola definitu predikatuak.....	21
2.3.6.- Emaitza aurkezteko beste era bat.....	21
3.- ANALIZATZAILEAREKIN LORTUTAKO EMAITZAK.....	22
3.1.- IZENEN, ADITZEN ETA ADJEKTIBOEN DEFINIZIOETARAKO PATROI-ARBOLAK ERATZEKO ERIZPIDEAK.....	22
3.2.- ANALISIAREN EMAITZAK ETA ARAZOAK.....	23
4.- BIBLIOGRAFIA.....	25
ERANSKINAK.....	27

## 1.- Sarrera

Lengoaia naturalaren prozesamenduaren (LNP) alorrean lexikoa geroz eta garrantzi handiagoa ari da hartzen. Linguistika konputazionalaren aldetik puntan dabilzan teorien arabera lexikoak izan behar du hizkuntzari buruzko jakitearen gordailu nagusia, aurreko teorian nagusitzen zen sintaxiaren kaltetan. Gauzak honela, LNPko sistemetako osagai nagusia bihurtu da lexikoa; horrek esan nahi du, besteak beste, izugarritzko informazio pila kodetu beharra dagoela LNPko sistema baten lexiko-sisteman. Kodeketa horretan lexiko-sarrera bakoitzari dagokion era guztietako informazioa zehaztu beharra dago, ortografikotik hasi eta morfologiko, sintaktiko eta semantikoraino. Maila errealeko sistema bat antolatu nahi denean berriz, bistakoa da kodetze-lan horri eskuz ekinez gero sekulako lana litzatekeela. Ahal den neurrian prozesua automatizatu beharra ikusten da, beraz.

Amsler-en lanetik (1980) hasi eta orain arte, lexikografia konputazionalaren alorrean ondoko galderei erantzuten saiatu dira ikerlariak: hiztegi arruntetan ez ote dago hitzei buruzko informazio asko eta asko? Hiztegi arruntetan bide dagoen jakintza lexiko ugari hori ezin al da egokitu LNPan erabilia izan dadin? Azken hamarkada honetan geroz eta errazago izan da hiztegi arruntak euskarri magnetikoan eskuratzea (Machine Readable Dictionary, MRD) eta horrek izugarri erraztu du hiztegi horien gaineko ikerketa.

Txosten honetan azaldu nahi dugun lana iker arlo horretan koka daiteke, hain zuzen ere. Hau da, hiztegi elebakar bateko definizioak hitzei —edo kontzeptuei— buruzko jakite semantikoaren iturri eta biltegitzat hartuz, definizio horiek analizatuz hiztegiko kontzeptuen arteko erlazio lexiko-semantikoak lortzea.

Horretarako tresna bat inplementatu eta erabili dugu: definizio-testuen analizatzaile bat; inplementazio horren nondik norakoak azaldu nahi genituzke hemen, ondoren aplikazioaren emaitzak azaldu eta komentatuz. Analizatzailea eroso erabili ahal izateko erraztasunak eta zehaztasunak ematea ere txosten honen helburuetako bat da. Inplementatutako tresna Alshawi-k (1987/1989) bere artikuluetan azaldutakoan oinarritu da.

Analizatzaile horrek pattern-matching hierarkikoan du funtsa, bere helburuetariko bat sendotasuna delarik. Sendotasun hori honela ulertu behar litzateke: nahiz eta analisi ez-osoak egin interpretazio semantiko partzial baina zuzenak lortzeko gaitasuna. Lehen esan dugun bezala, interesatzen zaigun interpretazio semantiko hori hiztegiko kontzeptuen arteko erlazio lexiko-semantikoetan oinarrituko da; hau da, kontzeptuen arteko