

MIATU Erraz

bilaketa liburutegia

Eskuliburua

2007ko apirila



ingeniaritza
ELEKA
Linguistikoa

ELEKA Ingeniaritza Linguistikoa, S.L.

www.eleka.net



Aurkibidea

1 Sarrera.....	3
1.1 Zer da Miatu?.....	3
1.2 Nori zuzenduta dago Miatu?.....	3
2 Miatu-ren instalazioa.....	3
2.1 Oinarrizko liburutegiaren instalazioa.....	3
WINDOWS.....	3
LINUX.....	3
2.2 Miatu Java.....	4
2.3 Miatu .NET.....	4
2.4 Miatu PHP.....	4
2.5 Miatu Python.....	4
3 Miatu-ren dependentziak.....	4
3.1 Oinarrizko dependentziak.....	5
Windows.....	5
Linux.....	5
3.2 Miatu Java.....	5
Windows.....	5
Linux.....	5
3.3 Miatu .NET.....	6
Windows.....	6
Linux.....	6
3.4 Miatu PHP.....	6
Windows.....	6
Linux.....	6
3.5 Miatu Python.....	6
Windows.....	6
Linux.....	6
4 Miatuaren erabilera.....	6
4.1 Miatuaren APIa.....	7
4.2 Miatuaren galdera lengoia.....	10
5 Laguntza teknikoa eta arazoaren konponketa.....	10



1 Sarrera

Atal honetan Miatu zer den, zein testuingurutan den erabilgarria eta nori zuzenduta dagoen azaltzen da.

1.1 Zer da Miatu?

Miatu euskarazko web guneak indexatzeko eta indize hauen gainean bilaketak egiteko funtzioak eskaintzen dituen liburutegi bat da. Miatuk euskararako bereziki garrantzitsuak eta erabilgarriak diren lematizazio teknika aurreratuak erabiltzen ditu bilaketak egiteko orduan emaitza hobekiak lortzeko. Bereziki diseinatuta dago erabiltzen erraza izateko, indexatzea egiten den bitartean bilaketak egin ahal izateko eta prozesu desberdinetatik aldi berean atzitu ahal izateko.

1.2 Nori zuzenduta dago Miatu?

Miatu ez dago erabiltzaile arruntei zuzenduta, web garatzaileei baizik. Miatu web garapeneko plataforma zabalduenetan integratzeko pentsatuta dago. Miaturi etekina atera ahal izateko beharrezkoa da web garapen inguruneko oinarrizko teknologien berri izatea.

2 Miatu-ren instalazioa

Miatu bi zati nagusitan banatzen da. Alde batetik, Miatu oinarrizko liburutegia, eta bestetik, liburutegi horretan oinarritzen diren goi mailako lengoaien interfazeak. Goi mailako lengoaien interfaze guztiek behar dute oinarrizko liburutegia instalatuta egotea, beraz, lehenengo eta behin oinarrizko liburutegia nola instalatu azalduko dugu labur-labur. Miatuaren instalazioa oso erraza da ikusiko duzuen moduan. Dena dela, badaude alde aurretik instalatu beharreko zenbait liburutegi. Hauei buruzko argibideak hurrengo atalean aurkituko dituzue.

2.1 Oinarrizko liburutegiaren instalazioa

WINDOWS

eleka_miatu.dll fitxategia PATH ingurune aldagaia adierazitako direktorioen barruan egon behar da. Horretarako bi aukera daude:

- eleka_miatu.dll C:\windows\ karpetan kopiatu (beste karpeta batzuk ere onargarriak izan daitezke).
- eleka_miatu.dll liburutegia zure gustuko karpeta batean kopiatu eta PATH inguruneko aldagaian karpeta hori gehitu.

LINUX

libeleka_miatu.so fitxategia estekatzailerak aurkitzeko moduko toki batean kopiatu behar da. Horretarako aukera desberdinak daude:

- libeleka_miatu.so /usr/local/lib karpetan kopiatu (beste karpeta batzuk ere onargarriak izan daitezke).
- libeleka_miatu.so zure gustuko karpeta batean kopiatu eta LD_LIBRARY_PATH inguruneko aldagaian balioan karpeta hori erantsi.



2.2 Miatu Java

Instalazio prozedura berbera da Linux eta Windowsentzat.

1. Lotura liburutegia (libmiatu.so linuxen eta miatu.dll windowsen) instalatu 2.1 atalean azaltzen den moduan.
2. miatu.jar fitxategia java inguruneak aurkitzeko moduko tokian jarri. Ingurunearen arabera modu desberdinean egiten da hori: konfigurazio fitxategi bidez, miatu.jar fitxategiaren bide osoa CLASSPATH inguruneke aldagaian gehituz, java komandoaren -cp parametroan miatu.jar gehituz,
...

2.3 Miatu .NET

Instalazio prozedura berbera da Linux eta Windowsentzat.

1. Lotura liburutegia (libeleka.so linuxen eta eleka.dll windowsen) instalatu 2.1 atalean azaltzen den moduan.
2. miatu.dll fitxategia .NET inguruneak aurkitzeko moduko tokian jarri. Ingurunearen arabera modu desberdinean egiten da hori: Visual Studio proiektuetan “Agregar Referencia” funtzioa erabiltzen da .NET assembly bat proiektuan gehitzeko. Mono modu oinarritzkoena mcs komandoaren -r parametroan gehitzea da.

2.4 Miatu PHP

Miatu PHP, PHPko “extension” bat da. Instalatzeko modu errazena php.ini fitxategia egokitzea da:

1. Sistemako php.ini fitxategian *extension_dir* aldagaiak esaten duen direktorioan kopiatu php_miatu.so fitxategia.
2. Sistemako php.ini fitxategian *extension = “php_miatu.so”* lerroa gehitu.

Miaturekin batera banatzen den php.ini fitxategian agertzen da prozedura honen adibide bat.

2.5 Miatu Python

eleka.py eta _eleka.py fitxategiak PYTHONPATH barruan sartu behar dira. Horretara bi aukera nagusi daude:

- eleka.py eta _eleka.py zure gustuko karpeta batean kopiatu eta PYTHONPATH aldagaian karpeta hori gehitu.
- eleka.py eta _eleka.py fitxategiak /usr/local/lib/python2.4/site-packages karpetan kopiatu (beste karpeta batzuk ere onargarriak izan daitezke).

3 Miatu-ren dependentsiak

Atal honetan Miatu instalatu ahal izateko kontutan hartu behar diren xehetasunak eta ase behar diren dependentsiak zeintzuk diren azaltzen da.



3.1 Oinarrizko dependentziak

Miatu C++en garatutako liburutegi bat da. Liburutegi honek hainbat oinarrizko dependentzia ditu. Oinarrizko dependentziak dira Miaturen interfaze guztiek (java, php, .NET...) behar dituzten dependentziak. Interfaze hauetako bakoitzak bere dependentzia gehigarri propioak ditu.

WINDOWS

Miatuk glib liburutegia (bere dependentziekin), Berkely DB 4.4, zlib eta pcre liburutegia behar ditu. Liburutegi hauek jarraian datozen tokietan lor ditzakezu behar izanez gero.

- glib, bere dependentzia diren gnulibconv eta gettext, eta zlib: <http://www.gimp.org/~tml/gimp/win32/downloads.html>
- Berkeley DB: <http://www.oracle.com/technology/software/products/berkeley-db/db/index.html> (4.4 bertsioa gomendatzen da).
- pcre: <http://www.psyon.org/projects/pcre-win32/>

LINUX

Miaturen linux bertsioak dituen dependentziak Windows bertsioaren berdina dira. Miatu Debian Etch makina baten garatu eta probatu da. Etch-en beharrezko paketeak aptitude bidez eskuratzea posible da. Hau da instalatuta egon behar diren paketeen zerrenda:

- libglib2.0 eta bere dependentziak.
- libdb4.4: Berkeley DB 4.4 bertsioa.
- libpcre3.
- Zlib1g.

Beste edozein Linux banaketatan ere egongo dira paketa hauek. Zure banaketaren eskuliburua kontsultatu.

3.2 Miatu Java

Miatu Java Miatu Javatik erabiltzea ahalbidetzen duen liburutegia da. Beharrezkoa da java instalatuta egotea erabiltzeko. Ez du oinarrizkoez aparte dependentzia gehigarrik.

WINDOWS

Sunen javaren 1.3 bertsioarekin dago estekatuta. Bertsio berriagoekin ere funtzionatuko du baina probatu gabe dago.

LINUX

Sunen javaren 1.5 bertsioarekin dago estekatuta.



3.3 Miatu .NET

Miatu .NET Miatu .NET plataformatik erabiltzea ahalbidetzen duen liburutegia da. Beharrezkoa da .NET instalatuta egotea erabiltzeko. Ez du oinarrizko ez parte dependentzia gehigarririk.

WINDOWS

MS .NET Framework 1.1ekin dago estekatuta.

LINUX

Mono 1.2rekin dago garatuta, Debian Etch-en datorren bertsioa.

3.4 Miatu PHP

Miatu PHP Miatu PHPtik erabiltzea ahalbidetzen duen liburutegia da. Beharrezkoa da PHP instalatuta egotea erabiltzeko. Ez du oinarrizko ez parte dependentzia gehigarririk.

WINDOWS

Windows plataformarako Miatu PHPren bertsiorik ez dago.

LINUX

PHP 4.3 bertsioekin dago konpilatuta.

3.5 Miatu Python

Miatu Python Miatu Pythonetik erabiltzea ahalbidetzen duen liburutegia da. Beharrezkoa da Python instalatuta egotea erabiltzeko. Ez du oinarrizko ez parte dependentzia gehigarririk.

WINDOWS

Windows plataformarako Miatu Pythonen bertsiorik ez dago.

LINUX

Python 2.4 bertsioekin dago konpilatuta.

4 Miaturen erabilera

Miaturen erabilera aztertzeko gomendagarriena liburutegiarekin batera banatzen diren adibideak aztertzea da. Atal honetan adibide guzti horiek amankomunean dituzten ezaugarriak eta kokapen finkorik ez duten beste zenbait xehetasun zerrendatzen dira.

Miatu liburutegi moduan aurkezten da. Liburutegi horrek 2 klase eskaintzen ditu. Erabiltzen den lengoaiaren arabera klase hauen eta beraien funtzioen deien sintaxia aldatu egiten da baina funtzionamendua mantendu egiten da. Jarraian bi klase horien funtzio nagusiak azaltzen dira labur C++ sintaxia erabiliz. Beste lengoaia guztietan funtzioen izenak, parametroen kopurua eta esanahia berdin-berdinak dira, sintaxia egokitu behar da besterik gabe. Lengoaia bakoitzari dagokion sintaxiaren xehetasunak liburutegiarekin batera banatzen diren



adibideetan daude dokumentatuta.

4.1 Miaturen APIa

Jarrian datorren taulan MIATUren interfaze publikoa osatzen duten funtzioak deskribatzen dira erraz kontsultatzeko formatuan. APIa berbera da lengoia guztientzat.

MIATU Erraz. Eskuliburua

Izena	Sarrerak	Azalpena
<code>class Miatu</code>	-	Miatu klasearen instantziak indizearen abstrakzioa dira. Nahi adina izan ditzakezu batera martxan, prozesu desberdinetatik ere bai eta Miatu arduratuko da beharrezkoa koordinazioaz. <code>Miatu</code> klasea <code>eleka::miatu</code> namespace barruan definitzen da.
<code>Miatu*</code> <code>Miatu::createInstance(const char *dbdir)</code>	<code>dbdir</code> : DB fitxategia (<code>miatu.index</code>) sortu behar duen direktorioaren izena. Existitzen den direktorio bat izan behar da. Bertan, indizeaz gain, sinkronizazio eta atzipen konkurrenterako beharrezkoak diren zenbait fitxategi sortuko dira.	Miatu klasearen instantzia berri bat. Direktorio horretan indize bat existitzen bada ireki egingo du eta bestela berri bat sortu.
<code>MiatuErrorCode</code> <code>Miatu::index(const char* url, const char* content, const char* mimetype)</code>	<ul style="list-style-type: none"> ● <code>url</code>: dokumentu honek izango duen izena (normalean <code>url</code> bat). ● <code>content</code>: dokumentuaren edukia. ● <code>mimetype</code>: edukiaren mime mota. "text/plain" edo "text/html" izan behar da. Content mota honekin bat etorri behar da. Baldintza hauek betetzen ez badira Miatu-ren funtzionamendua ez dago definituta. 	<code>MEC_SUCCESS</code> dokumentua ondo indexatu bada, errore kode bat bestela.
<code>void Miatu::close()</code>	-	Indizera erreferentzia askatzen du. Lengoia gehienetan ez dago deitu beharrik. Salbuespena Java da. Erreferentzia guztiak askatzen indize datu basea itxi egiten da.
<code>ResultIterator*</code> <code>Miatu::lookup(const char *query)</code>	<code>query</code> : ebatzi beharreko galdera.	Miatuk galdera sintaxi sinple baina ahaltsua eskaintzen du, kasu gehienetan nahikoa izango dena. Sintaxiaren nondik norakoak eskuliburuko 4.2 atalean azaltzen dira. Eraitza <code>ResultIterator</code> instantzia bat da edo <code>NULL</code> erroreren bat gertatu bada.



MIATU Erraz. Eskuliburua

Izena	Sarrerak	Azalpena
<code>class ResultIterator</code>	-	Emitza moduan URL multzo baten abstrakzioa den klasearen instantzia bat ematen du. Instantzia honek indizera erreferentzia bat gordetzen duenez, posible da <code>Miatu::close</code> deitu eta gero ere instantzia erabiltzea. <code>ResultIterator</code> klasea <code>eleka::miatu</code> namespace barruan definitzen da.
<code>const char * ResultIterator::next()</code>	-	Emitza multzoko hurrengo osagaia ematen du edo NULL, osagai gehiago ez badaude.
<code>void ResultIterator::close()</code>	-	Indizera erreferentzia askatzen du. Lengoia gehienetan ez dago deitu beharrik. Salbuespena Java da. Erreferentzia guztiak askatzen indize datu basea itxi egiten da.



4.2 Miaturen galdera lengoia

Miaturen galdera lengoia sinplea baina ahaltsua da eta kasu gehienetan nahikoa izatea espero da.

Miaturen galdera lengoia lookup funtzioaren deietan agertzen da. Karaktere alfabetikoz osatutako hitzak bakarrik onartzen ditu. Emaiza moduan galderan emandako hitz guztiak dituzten dokumentuen zerrenda ematen du, hau da, hitz guztien arteko AND bat egiten du.

Indizean bilaketa egin aurretik hitzak lematizatu egiten ditu. Honek ematen dizkio Miaturi bere ezaugarri bereziak, Euskararako ezinbestekoak direnak. Adibidez, bilaketa gako moduan “amaren” emanaz gero, indizean “amaren” bilatzeaz gain “ama”, “amak”, “amek”, “amaren”, “amatik”, ... bilatuko ditu, baina ez “amonaren” edo “amigdala”, MIATUk dituen lematizazio funtzio aurreratuak ez dituen bilatzaile batek egingo lukeen moduan.

5 Laguntza teknikoa eta arazoen konponketa

Liburutegi honen erabileran arazoren bat aurkituz gero jo *ELEKA Ingeniaritza Linguistikoa, S.L.*ra. kontaktu informazioa aurkitzeko tokirik egokiena www.eleka.net web gunea da.