

MIATU Erraz

bilaketa liburutegia

Eskuliburua

2009ko abendua



ingeniaritza
ELEKA
Linguistikoa

ELEKA Ingeniaritza Linguistikoa, SL

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
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.....	5
3.1 Oinarrizko dependentziak.....	5
Linux.....	5
3.2 Miatu Java.....	5
Linux.....	5
3.3 Miatu .NET.....	5
Linux.....	5
3.4 Miatu PHP.....	5
Linux.....	6
3.5 Miatu Python.....	6
Linux.....	6
4 Miatuaren erabilera.....	6
4.1 Miatuaren APIa.....	6
4.2 Miatuaren galdera lengoia.....	9
5 Laguntza teknikoa eta arazoaren konponketa.....	9



1 Sarrera

Atal honetan Miatu zer den, zein testuinguruan 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 hobeak 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. Miatu etekina atera ahal izateko beharrezkoa da web garapen ingurune oinarritzko teknologien berri izatea.

2 Miatu-ren instalazioa

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

2.1 Oinarritzko liburutegiaren instalazioa

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 ingurune aldagairen balioan karpeta hori erantsi.



2.2 Miatu Java

Instalazio prozedura berbera da Linux 32 bit eta Linux 64 bitentzat.

1. Lotura liburutegia (libmiatu.so) 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 32 bit eta Linux 64 bitentzat.

1. Lotura liburutegia (libeleka.so) 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 oinarrikoena 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 kopia php_miatu.so fitxategia.
2. Sistemako php.ini fitxategian *extension = “php_miatu.so”* lerroa gehitu.

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



3 Miatu-ren dependentziak

Atal honetan Miatu instalatu ahal izateko kontutan hartu behar diren xehetasunak eta ase behar diren dependentziak 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.

LINUX

Miaturen linux bertsioak dituen dependentziak berdinak dira 32 nahiz 64 biteko bertsioetarako. Miaturen 2009ko abenduako bertsioa Ubuntu 9.04 makina baten garatu eta probatu da. Hau da instalatuta egon behar diren paketeen zerrenda:

- libglib2.0 eta bere dependentziak.
- libdb4.7: Berkeley DB 4.7 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 gehigarririk.

LINUX

Sunen javaren 1.6 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 oinarrizkoez aparte dependentzia gehigarririk.

LINUX

Mono 2.0rekin dago garatuta, Ubuntu 9.04rekin datorren bertsioa.

3.4 Miatu PHP

Miatu PHP Miatu PHPtik erabiltzea ahalbidetzen duen liburutegia da.



MIATU Erraz. Eskuliburua ELEKA Ingeniaritza Linguistikoa, SL
Beharrezkoa da PHP instalatuta egotea erabiltzeko. Ez du oinarritzko ez aparte dependentzia gehigarrik.

LINUX

PHP 5.2.6 bertsioarekin dago konpilatuta.

3.5 Miatu Python

Miatu Python Miatu Pythonetik erabiltzea ahalbidetzen duen liburutegia da. Beharrezkoa da Python instalatuta egotea erabiltzeko. Ez du oinarritzko ez aparte dependentzia gehigarrik.

LINUX

Python 2.6 bertsioarekin 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. Lengoia bakoitzari dagokion sintaxiaren xehetasunak liburutegiarekin batera banatzen diren adibideetan daude dokumentatuta.

4.1 Miaturen APIa

Jarraian 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. Miatu klasea <code>eleka::miatu</code> namespace barruan definitzen da.
<code>Miatu*</code> <code>Miatu::createInstance(const char *dbdir)</code>	dbdir: 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">● url: dokumentu honek izango duen izena (normalean url bat).● content: dokumentuaren edukia.● mimetype: 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.	MEC_SUCCESS 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.



MIATU Erraz. Eskuliburua

Izena	Sarrerak	Azalpena
ResultIterator* Miatu::lookup(const char *query)	query: ebatzi beharreko galdera.	Miatuk galdera sintaxi simple baina ahaltsua eskaintzen du, kasu gehienetan nahikoa izango dena. Sintaxiaren nondik norakoak eskuliburuko 4.2 atalean azaltzen dira. Eraitza ResultIterator instantzia bat da edo NULL errorearen bat gertatu bada.
class ResultIterator	-	Eraitza moduan URL multzo baten abstrakzioa den klasearen instantzia bat ematen du. Instantzia honek indizera erreferentzia bat gordetzen duenez, posible da Miatu::close deitu eta gero ere instantzia erabiltzea. ResultIterator klasea eleka::miatu namespace barruan definitzen da.
const char * ResultIterator::next()	-	Eraitza multzoko hurrengo osagaia ematen du edo NULL, osagai gehiago ez badaude.
void ResultIterator::close()	-	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, SL*ra. kontaktu informazioa aurkitzeko tokirik egokiena www.eleka.net web gunea da.