

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Informatika Ingeniaritzako Gradua

Software Ingeniaritza

Gradu Amaierako Proiektua

Oso corpus handien indexazioa, IR teknikak erabiliz

Egilea

Maialen Otamendi Zurutuza

informatika
fakultatea



facultad de
informática

2016

Laburpena

Proiektu honetan testu bilduma oso handien gainean (corpusak) kontsulta aurreratuak egiteko web-aplikazio bat garatu da. Testuak informazio linguistikoaz aberastuta daude, eta kontsultak ezaugarri linguistiko horietan oinarritu dira. Arreta handia jarri zaio aplikazioaren eraginkortasunari, hots, emaitzak jasotzeko denbora laburra behar izateari. Informazioaren berreskurapenean (IR, Information Retrieval) erabili den tresna *Lucenen* oinarritutako *Elasticsearch* da.

Gaien aurkibidea

Laburpena	iii
Gaien aurkibidea	v
Irudien aurkibidea	ix
Taulen aurkibidea	xi
1 Sarrera	1
2 Proiektuaren kudeaketa-plana	3
2.1 Irismena	3
2.1.1 Betekizunak	3
2.1.2 Mugarriak	4
2.1.3 Irismen-mailak	5
2.1.4 Irismenaren kudeaketa	5
2.1.5 Lanaren Deskonposaketa Egitura	5
2.2 Emangarri nagusiak	6
2.3 Eskuraketak	6
2.4 Denboraren kudeaketa	8
2.4.1 Atazen definizioa	8

2.4.2	Gantt diagrama	8
2.4.3	Estimazioak	11
2.5	Kalitate-plana	12
2.5.1	Kalitatearen kontrola	13
2.6	Arriskuak	13
2.7	Komunikazio-plana	14
2.7.1	Interesatuak	14
2.7.2	Komunikazio-plana	15
3	Aurrekariak	17
3.1	Testuingurua	17
3.1.1	Anotazio-amarauna	17
3.1.2	Analisi linguistikoak	18
3.2	Armiarma	18
3.2.1	Arkitektura	19
3.3	Proiektu honi buruz	20
3.3.1	Elasticsearch	21
3.3.2	Datu-base erlazionalekiko desberdintasunak	24
3.3.3	Funtzionamendua	24
4	Diseinua	27
4.1	Elasticsearch	27
4.1.1	Dokumentuen eta analisisien diseinua	27
4.1.2	<i>Shard</i> eta <i>replica</i> kopuruak zehaztu	30
4.1.3	Datu-egituraren diseinua	30
4.1.4	Bilaketak	32
4.2	Web-aplikazioaren interfazea	33
4.2.1	Formulario dinamikoa	37

5	Proiektuaren garapena	41
5.1	Datuen formatua aldatu	41
5.2	Datuen indexazioa	43
5.3	Elasticsearch-PHP konfigurazioa	43
5.4	Web-aplikazioaren ezaugarri eta erabilpen-kasuak	44
5.4.1	Aplikazioaren egitura eta itxura	45
5.4.2	Formularioa	46
5.4.3	<i>Bilaketa egin/emaitzak ikusi</i> erabilpen-kasua	47
5.4.4	<i>Analisiak ikusi</i> erabilpen-kasua	50
5.4.5	<i>Orriz aldatu</i> erabilpen-kasua	51
5.5	Erabilitako teknologiak	52
6	Esperimentuak	53
7	Jarraipena eta Kontrola	57
7.1	Desbideraketak	57
7.2	Gantt diagrama	60
7.3	Mugarriak	60
7.4	Irismen-maila	60
7.5	Kalitate-plana	62
7.6	Eskuraketak	62
7.7	Arriskuak	62
8	Ondorioak	63
8.1	Ikasitako lezioak	64
8.2	Etorkizuneko lanak	64

Irudien aurkibidea

2.1	LDE diagrama.	7
2.2	Gantt diagrama.	10
3.1	Hitzen eta analisisien arteko erlazioa datu-baseetan.	19
3.2	Armiarmaren arkitektura. Iturria: Jagoba Gascón Sánchezen GAPa, 2013.	20
3.3	<i>Shard</i> bat eta bi <i>Sharden</i> arteko diferentzia.	22
3.4	<i>Elasticsearch</i> en kontzeptuak ulertzeko irudia.	23
3.5	Nodoren bat erortzen bada zer gertatzen den erakusten duen irudia.	23
3.6	<i>Index</i> APIaren funtzionamendua	25
3.7	<i>Get</i> APIaren funtzionamendua	26
4.1	Dokumentu bateko hitz baten adibidea.	29
4.2	Guk definitutako <i>shard</i> eta <i>replica</i> kopuruak.	30
4.3	Guk definitutako datu-egitura edo <i>mappinga</i> .	31
4.4	Hitz bat bilatzeko eskaeraren adibidea.	33
4.5	Bi hitz bilatzeko eskaeraren adibidea.	34
4.6	Hiru hitz bilatzeko eskaeraren adibidea.	34
4.7	Web-aplikazioaren interfazea: "Baina"lemaren bilaketa.	35
4.8	Web-aplikazioaren interfazea 4.7 irudiko bilaketaren emaitzekin.	35
4.9	Web-aplikazioaren interfazea hitz baten analisiarekin.	36

4.10	Interfazeko formularioaren funtzionamendua.	37
4.11	<i>Izena</i> kategoriaren ezaugarriak.	38
4.12	Formularioan ezaugarri lingustikoen izen osoak agertzen dira.	39
5.1	Dokumentuek izan behar duten egitura zehatza.	42
5.2	Dokumentu baten egitura.	42
5.3	Dokumentuak indexatzeko exekutatu beharreko agindua.	43
5.4	Elasticsearch dependentzia.	44
5.5	Bezeroa sortzen.	44
5.6	Web-aplikazioaren interfazearen egitura (HTML kodea).	45
5.7	Formularioaren sorkuntza (sasi-kodea).	46
5.8	<i>Bilaketa egin/emaitzak ikusi</i> erabilpen-kasuaren sekuentzia-diagrama. . .	47
5.9	Elasticsearch-PHP bilaketa baten adibidea.	48
5.10	Emaitzak nola pantailaratu erakusten duen kode zatia.	49
5.11	Analisiak ikusi erabilpen-kasuaren sekuentzia diagrama.	50
5.12	Analisiak (Ajax kodea).	51
5.13	<i>analisia.php</i> orrian analisisia nola jaso.	51
5.14	Orrialde aldaketa (JavaScript kodea).	51
6.1	Aplikazioaren erantzun-denborak dokumentu kopuruaren arabera.	55
7.1	Gantt diagrama.	61

Taulen aurkibidea

2.1	Atazen definizioa.	9
2.2	Denboraren estimazioa.	11
2.3	Arriskuak.	14
2.4	Interesatuak.	14
3.1	Datu-base erlazionalekiko desberdintasunak.	24
6.1	Aplikazioaren erantzun-denborak eta emaitza kopuruak.	54
7.1	Desbideraketak.	58

1. KAPITULUA

Sarrera

Dokumentu hau 2015-2016 ikasturtean Maialen Otamendi Zurutuzak egindako Gradu Amaierako Proiektuari dagokio. Proiektuaren zuzendariak Xabier Artola Zubillaga eta Aitor Soroa Etxabe, UPV/EHUko Donostiako Informatika Fakultateko irakasleak, izan dira.

Proiektua IXA taldekoek proposatutakoa da. IXA taldea Euskal Herriko Unibertsitateko Donostiako Informatika Fakultatean kokatzen den ikerketa talde bat da. Informatikariz eta hizkuntzalariz osatuta dago eta Hizkuntza Naturalaren Prozesamenduaren arloa jorratzen du.

Proiektu honetan corpusen¹ gainean kontsulta aurrerratuak egiteko web-aplikazio bat garatu da. Aplikazio hau Euskaltzaindiaren Lexikoaren Behatokia egitasmoan kokatzen da, eta bere helburu nagusia corpus ezberdinekin osatutako anotazio-amaraun konplexu baten gainean hizkuntza-kontsultak egiteko aukera ematea da.

Horretarako, Armiarma izeneko beste web-aplikazio batean oinarritu gara. Anotazio linguistiko konplexuak kudeatzeko aplikazioa da. Datu-base erlazionaletan eta XML datu-baseetan oinarritzen da, eta bere gabeziak erakutsi ditu. Testu bilduma oso handien gainean kontsulta aurrerratuak egin nahi direnean motel erantzuten du, eta, beraz, erantzundebora hori jaisteko beste teknologia batekin probatzea izan da helburua. Proiektu honetan informazioaren berreskurapena (ingelesez, *Information Retrieval* edo *IR*) teknologia erabili da, *Elasticsearch* tresnari esker.

¹Corpusak testu bilduma handiak dira.

Proiektuak, beraz, bi atal nagusi ditu:

- *Elasticsearch* tresnaren ikerketa.
- Corpus anotatua (informazio linguistikoz hornitua) errepresentatzea, *Elasticsearch*en bidez indexatu eta kontsultatu ahal izateko .
- Web-aplikazioaren garapena.

2. KAPITULUA

Proiektuaren kudeaketa-plana

Kapitulu honetan proiektuaren irismena, lan-metodologia, emangarriak, baliabideak, garapenaren planifikazioa, kalitate-plana eta arriskuen plana azaltzen dira. Proiektuaren garapenerako erabiliko den tresna berria denez (bai zuzendarientzat eta baita ikaslearentzat ere), litekeena da proiektuan zehar planifikazioa aldatu behar izatea.

2.1 Irismena

Atal honetan proiektuaren irismena zein den azalduko da.

Hasteko, argi geratu behar da helburua zein den zehazki. Web-aplikazio bat garatu behar-ko da informazioaren berreskurapenerako, *Elasticsearch* tresna erabiliz, eta Armiarma aplikazioaren erantzun-denbora murrizten saiatuko gara.

Jarraian betekizunak, mugak, irismen-maila, irismenaren kudeaketa eta lanaren deskonposaketa-egitura definitu dira.

2.1.1 Betekizunak

Proiektuan zehar Armiarman oinarritutako beste web-aplikazio bat garatuko da. Web-aplikazioak honako betekizunak izan behar ditu:

- Web-aplikazioak hiru atal nagusi izango ditu: beti ikusgai egongo den formularioa,

bilaketen emaitzak pantailaratzeko atala eta bilatutako hitzen analisiak pantailaratzeko beste atal bat.

- Bilaketa-interfazeak dinamikoa izan behar du neurri handi batean. Beraz, fitxategi bat sortuko da analisi linguistikoaren eskema definitzeko. Horretarako, Armiamaren garapenean sortu zuten Relax NG eskeman oinarrituko gara, baina ez da beharrezkoa izango ezaugarri guztiak sartzea, ez baita hori proiektuaren helburua. Ezaugarrien aukeraketa bilera batean egingo da.
- Bilaketa-interfazeak ondorengo ezaugarriak bete behar ditu:
 - Formularioko eduki guztietan bakoitzaren izena osorik ikusi beharko da.
 - Hiru hitz bilatzeko aukera eman behar du, bakoitza ezaugarri linguistikoen arabera.
 - Bigarren eta hirugarren hitzen posizioak hautatzeko aukera eman behar du (aurretik edo atzetik, eta zenbat pauso aukeratzeko aukera). Posizioa aukeratzekoan lehenengo hitza hartuko da beti oinarritzat.
 - Analisi zuzena edo ez-zuzena hautatzeko aukera eman behar du.
- Informazioaren berreskurapenerako Elasticsearch tresna erabiliko da.

2.1.2 Mugarriak

Proiektuak dituen atazak burutzeko egokia da epe-muga batzuk ezartzea. Ez da beharrezkoa izango, inondik inora, epe hauek zehatz-mehatz betetzea, baina proiektuaren garapenerako lagungarria izango da.

Hori azalduta, jarraian mugarriak definituko dira:

- **2014-12-26:** *Elasticsearch*i buruzko ikerkuntza bukatu.
- **2015-03-30:** Web-aplikazioaren diseinua bukatu.
- **2015-07-03:** Web-aplikazioaren garapena bukatu.
- **2015-08-31:** Memoria bukatu.

2.1.3 Irismen-mailak

Irismena definitzeko bi maila ezarriko ditugu: oinarritzkoa eta aurreratua. Oinarritzkoa iristea beharrezkoa izango da, eta maila aurreratura iristea, berriz, ez. Definitutako mugariak eta ordu kopurua kontuan hartuta erabakiko da maila honetara iritsiko garen edo ez. Atzerapenak gertatzen hasten badira edo planifikatutako ordu kopuruak gainditzen hasten bagara, oinarritzko mailan geratuko gara.

1. Oinarritzko maila: Aurrez definitutako betekizunak betetzeaz gain, interfazean gerta daitezkeen erroreak zuzendu beharko dira. Armiarma aplikazioarekiko eraginkortasuna hobetzen ere siatuko gara.
2. Maila aurreratua: Oinarritzko mailan zehaztutakoa betetzeaz gain, web-aplikazioari estatistikak erakusteko atal bat gehituko zaio. Estatistikak egin den bilaketari buruzkoak izango dira, kategorietan oinarritutako portzentaiak adieraziz. Horretaz aparte, erabiltzaileari begira interfaze erraz, intuitibo eta txukuna garatuko da.

2.1.4 Irismenaren kudeaketa

Gerta daiteke irismenean definitutako atalen bat aldatu behar izatea, aurreikusi ez den arazoren bat gertatu delako edo atzerapenak direla medio. Kasu horretan, litekeena da mugariak aldatu beharra izatea, gutxienez proiektua oinarritzko mailara irits dadin. Proiektuaren betebeharrak betetzeak beste ezeren aurretik egon behar du beti.

Erabaki horiek guztiak, hala ere, ez dira edonola hartuko. Bilera batean aztertu eta erabakiko da, tutoreen eta ikaslearen artean adostasun batera iritsiz.

2.1.5 Lanaren Deskonposaketa Egitura

Lanaren deskonposaketa-egiturak proiektua atal eta azpiatazetan banatzen ditu. Kasu honetan, proiektua bost ataletan banatu dugu: kudeaketa, ikerkuntza, diseinua, garapena eta dokumentazioa.

- **Kudeaketa:** Kudeaketaren barruan plangintza, jarraipena eta kontrola sartzen dira. Plangintzan, garapena planifikatzeaz gain, irismena, atazak eta estimazioak definitzen dira. Jarraipen eta kontrola proiektu osoan zehar egingo da. Lehenengoa, batez

ere, bileren bidez. Bigarrena, berriz, arriskuak, kalitatea eta gertatutako aldaketak kontrolatuz eta eguneratuz.

- **Ikerkuntza:** *Elasticsearchi* buruz gauza gutxi dakigunez, beharrezkoa izango da ikerkuntzari denbora eskaintzea, hori eginez gero, ziurrenik, etorkizunean (web-aplikazioa garatzerakoan, batez ere) denbora murritzuko baitugu.
- **Diseinua:** Lehenik eta behin, datu-egituraren eskema espezifikatuko da eta, ondoren, web-aplikazioaren interfazea diseinatu beharko da. Datu-egitura kontu handiz sortuko da, errendimendua hobetzen saiatu behar gara eta.
- **Garapena:** Garapena ez da web-aplikazioa garatzea soilik. Ezer baino lehen lan-ingurunea prestatu beharko da (behar den softwarea instalatu), beharrezko script-ak programatu (tutoreen bidez eskuratutako datuak formatuz aldatu beharko dira ziurrenik) eta, amaitzeko, behar diren proba guztiak egin beharko dira aplikazioa behar den bezala dabilela ziurtatzeko.
- **Dokumentazioa:** Azkenik, memoria idatzi beharko da eta aurkezpena prestatzeko materiala ere bai.

[2.1](#) irudiak aurreko guztia diagrama moduan biltzen du.

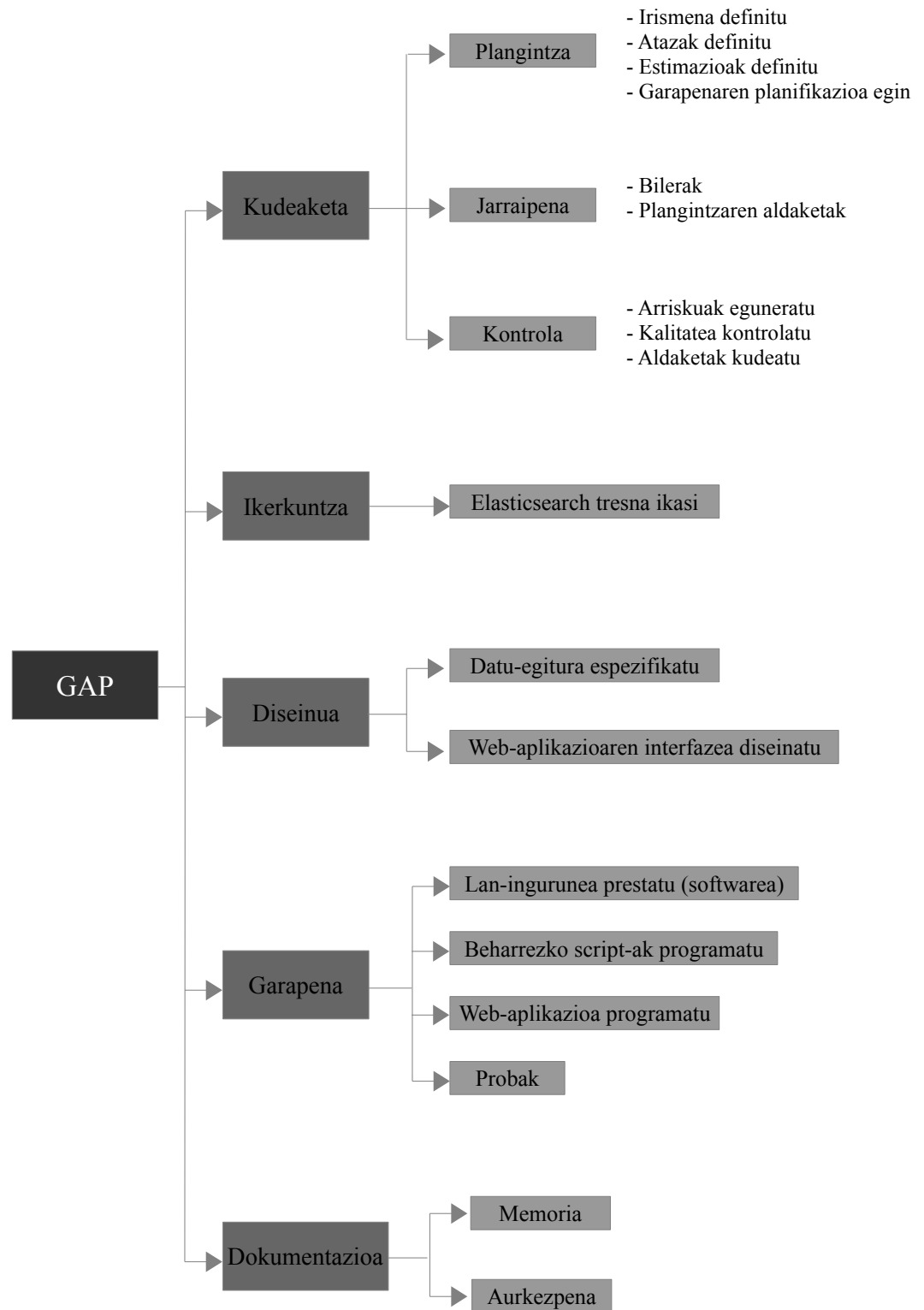
2.2 Emangarri nagusiak

Proiektuak hiru emangarri nagusi izango ditu:

- Web-aplikazioaren kodea: Memoriarekin batera igo beharko da ADDI plataformara.
- Memoria.
- Defentsa-egunean erabiliko den aurkezpen-dokumentua.

2.3 Eskuraketak

Proiektu honetan eskuratu beharko diren baliabideak software motakoak izango dira (ikasleak erabiliko duen konputagailu pertsonalaz aparte), eta, gehienak software-libreak direnez, ez da inongo gasturik egongo. Hala ere, softwareaz aparte, *Elasticsearchi* buruzko



ikasketa egiteko zuzendariek EHUko liburutegian liburu bat eskatu dute ikasleari baliagarri izango zaiolakoan.

Gauzak horrela, informazioaren berreskurapenerako Elasticsearch tresna erabiliko denez, hori instalatu beharko da lehendabizi. Behar diren script-ak programatzeko Eclipse ingurunea erabiliko da, eta web-aplikazioa programatzeko Sublime Text testu-editorea erabiliko da. Azken hori ez da guztiz libre, baina lizentzia ez edukitzeak ez du mugarik ezartzen, noizbehinka alerta moduko mezua bat agertzen da soilik. Beraz, lizentziarik gabe erabiliko da.

Web-aplikazioa martxan jartzeko zerbitzari lokala erabiliko da, MAMP aplikazioa erabiliz.

Horrez gain, Elasticsearch erabiltzeko gehienbat, litekeena da horrekin batera dependentsia duen zerbait instalatu beharra izatea.

2.4 Denboraren kudeaketa

Atal honetan denboraren kudeaketa egingo da. Lehendabizi atazak definitu beharko dira eta, ondoren, denboren estimazioa egin.

2.4.1 Atazen definizioa

Proiektu bat garatu aurretik beharrezkoa da lana atazetan banatzea, planifikazioa ahalik eta zehatzena izatea lortu dadin. Hori dela eta, [2.1](#) taulan ikus daitezke proiektu honetarako zehaztu diren atazak eta bakoitzaren azalpena.

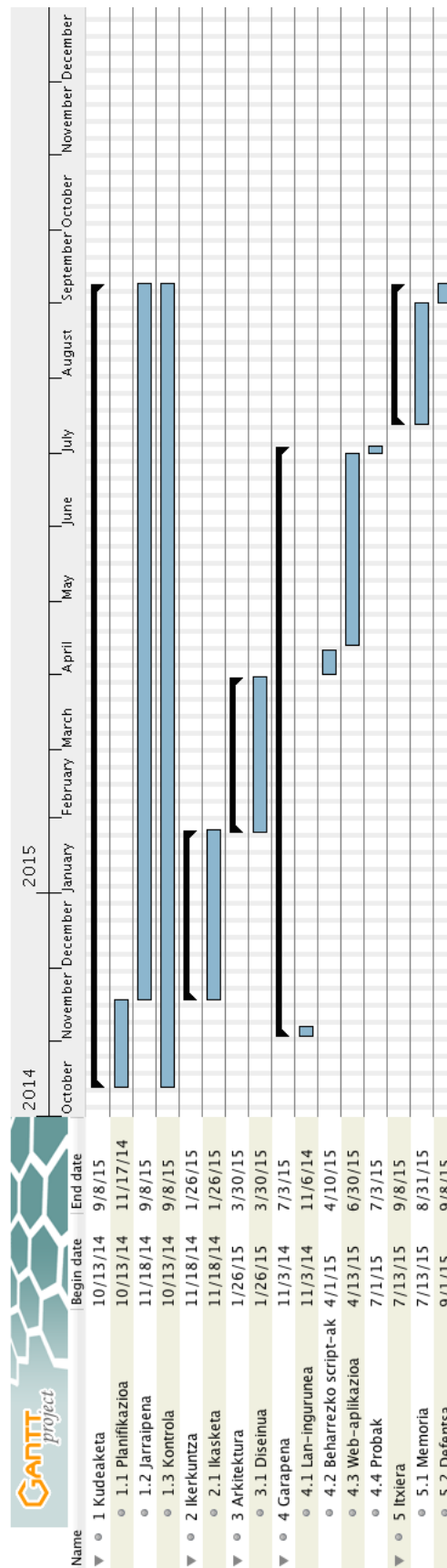
2.4.2 Gantt diagrama

Gantt diagramak aurrez definitutako atazak zenboran zehar noiz beteko diren zehazten du. [2.2](#) irudian agertzen da proiektu honetarako planifikatu den egutegia.

Esan beharra dago, urtarrean, azterketak direla-eta, ordu gutxi sartuko direla.

ID	Ataza	Deskribapena
1. Kudeaketa		
1.1	Plangintza	Proiektu osoaren planifikaizoa: irismena zehaztu, atazak definitu, emangarriak identifikatu, denbora kudeatu, kalitate-plana eta arriskuak identifikatu.
1.2	Jarraipena	Lanaren jarraipena egiteko bilerak egingo dira tutoreen eta ikaslearen artean. Plangintzaren edozein aldaketa aktetan jasoko da.
1.3	Kontrola	Arriskuak, kalitatea eta edozein aldaketa kudeatu beharko dira.
2. Ikerkuntza		
2.1	Elasticsearch	Tresna hau proiektuaren interesatuentzat berria denez, informazioa bilatu eta honi buruz ikasi.
3. Diseinua		
3.1	Datu-egitura	Garatzen hasi baino lehen, datuen egitura espezifikatu behar da.
3.2	Web-aplikazioaren interfazea	Web-aplikazioaren interfazea ere diseinatu beharko da interesatuen artean.
4. Garapena		
4.1	Lan-ingurunea	Proiektu honen kasuan, beharrezko softwarea deskargatu beharko da ezertan hasi aurretik.
4.2	Beharrezko Script-ak	Datu-egitura behar den formatura pasatzeko script-en bat programatu beharko da.
4.3	Web-aplikazioa	Proiektu honen funtsa, web-aplikazioa garatzea <i>Elasticsearch</i> en funtzionamendua ikusteko.
4.4	Probak	Hainbat proba egingo dira, bilaketa mota ezberdinen bidez, proiektua bukatutzat eman aurretik.
5. Dokumentazioa		
5.1	Memoria	Proiektuaren gorabeherak azaltzen dituen memoria idatziko da.
5.2	Aurkezpena	Defentsaren egunerako aurkezpen bat prestatu beharko da.

2.1 Taula: Atazen definizioa.



2.2 Irudia: Gantt diagrama.

2.4.3 Estimazioak

Denbora kudeatzeko azken pausoa ataza bakoitzaren estimazioa egitea da, hau da, ordu kopuruen gutxi gorabeherako bat ezartzea. Ez da beharrezkoa izango zehatz-mehatz he-men jarritakoa betetzea, baina ondo legoke, ahal den heinean, bertatik asko ez urruntzea. [2.2](#) taulan ikus daitezke planifikatutako orduak.

ID	Ataza	Estimazioa
Kudeaketa		
1.1	Plangintza	25 ordu
1.2	Jarraipena	13 ordu
1.3	Kontrola	14 ordu
Kudeaketa: 52 ordu		
Ikerkuntza		
2.1	Elasticsearch	60 ordu
Ikerkuntza: 60 ordu		
Diseinua		
3.1	Datu-egitura	90 ordu
3.2	Web-aplikazioaren interfazea	2 ordu
Diseinua: 92 ordu		
Garapena		
4.1	Lan-ingurunea	6 ordu
4.2	Beharrezko script-ak	15 ordu
4.3	Web-aplikazioa	65 ordu
4.4	Probak	2 ordu
Garapena: 88 ordu		
Dokumentazioa		
5.1	Memoria	92 ordu
5.2	Aurkezpena	8 ordu
Dokumentazioa: 100 ordu		
GUZTIRA: 392 ordu		

2.2 Taula: Denboraren estimazioa.

- **Kudeaketa:** Kudeaketa garrantzia handia duen atala da, baina ordu kopuruei dago-kienez, proportzioan, denbora gutxien eramango duena. Lan honetan pisu gehien duen ataza, zalantzarik gabe, planifikazioa da. Arreta handia jarri behar zaio honi, izan ere, dena ondo loturik uztea lortzen bada, etorkizunean buruhauste asko ekidin ditzakegu. Bestalde, jarraipena eta kontrola proiektu osoan zehar egingo da tarteka; beraz, 13-14 ordu estimatu dira.

- **Ikerkuntza:** Ez da ahaztu behar proiektu honen interesatuak ez dutela Elasticsearch ezagutzen, eta, beraz, ikasleak horri buruz ikasten denbora pasatu beharko du ahalik eta datu-egituraren diseinu egokiena lor dezan. 50 ordu planifikatu dira guztira.
- **Diseinua:** Behin ikerkuntza-atala burututa, aplikazioaren eta datu-egituraren diseinua egitera pasatu beharko gara. Arreta berezia jarriko zaio atal honi, aplikazioaren errendimendua jokoan dago eta. Horregatik, 95 ordu planifikatu dira.
- **Garapena:** Garapenean ez da web-aplikazioa bakarrik sartzeko, nahiz eta hori izan denbora gehien eramango duen ataza. Lan-ingurunearen prestakuntzari ere eskaini beharko zaio denbora, behar diren script-ei, eta amaierako probei ere bai. Gauzak horrela, 88 ordu estimatu dira guztira.
- **Dokumentazioa:** Atal honetan denbora gehien iraungo duen ataza memoria izango da. Proiektu osoan zehar izandako gorabeherak azaldu behar direnez, 92 ordu igarotzea espero da. Defentsa eguneko aurkezpena prestatzeko, berriz, ez da hainbeste denbora beharko (8 orduko estimazioa).

2.5 Kalitate-plana

Proiektuaren kalitatea planifikatzea oso garrantzitsua da, izan ere, bezeroari begira ez da nahikoa betekizun eta murriztapenak betetzea. Gainera, bezeroaren eskakizun maila gero eta handiagoa izan ohi da. Jarraian, proiektua kalitatezkoa izateko bete behar diren irizpideak azaltzen dira:

- **Denboraren kalitatea:** Proiektua ordu askokoa da, eta ezinbestekoa izango da denbora ahalik eta eraginkorren erabiltzea.
- **Emaitzaren kalitatea:** Burutuko den web-aplikazioaren kalitatea ere nolabait kontrolatu beharko da.
- **Kodearen kalitatea:** Aplikazioaren kodea txukuna izateak garrantzia handia du, batez ere etorkizunera begira.
- **Dokumentazioaren kalitatea:** Proiektua amaitzeko idatzi behar den memoria eta aurkezpenen kalitatea ere bermatu beharko da.

2.5.1 Kalitatearen kontrola

Aurreko irizpideak definitu ondoren, atal bakoitzaren kalitatea nola kontrolatuko den azaldu behar da.

- **Denboraren kalitatea:** Denbora ahalik eta gehien aprobetxatzeko lantokia finkatzea ezinbestekoa da. Hortaz, gradu-amaierako proiektua egiteko Informatika Fakultatean dagoen gelan arituko da lanean ikaslea.
- **Emaitzaren kalitatea:** Web-aplikazioak badu bere zailtasuna, eta interfazeak, besteak beste, erabiltzailearentzat erabilerraza izan behar du. Bestalde, aplikazioaren errendimendua hobetzen ere saiatuko gara, hau da, bilaketa bat egiteko behar den denbora murrizten.
- **Kodearen kalitatea:** Ez dakigu kodea norbaitek berrerabiliko duen; beraz, badaezpada ere kodea txukun mantendu beharko da.
- **Dokumentazioaren kalitatea:** Dokumentazioak izugarrizko garrantzia du, eta, beraz, arreta berezia jarriko zaio atal honi. Ikasleak ez du zerotik hasi beharko hau egitean, izan ere, Informatika Fakultateko web-orrian eskuragarri dago memoria egiteko txantiloia.

2.6 Arriskuak

Proiektu bat aurrera eramateko garaian beti sartzen dira arriskuak jokoan, eta hori nolabait kontrolatu beharra dago. Gerta daitezkeen arazoei aurre egiteko gai izan behar dugu eta horretarako ezinbestekoa da arriskuak kudeatzea.

Lehenengo pausoa arrisku horiek identifikatzea izango da, eta, ondoren, arrisku horiei aurre egiteko zein neurri hartuko diren azalduko da. [2.3](#) taulan ikus daiteke hori guztia laburtuta.

Hauek dira, [2.3](#) taulan agertzen diren arriskuren baten aurrean egonez gero, hartuko diren neurriak:

- **A1:** Web-aplikazioaren kodearekin oso kontuz ibili beharko gara. Horretarako, edozein aldaketa garrantzitsu egiten bada, memento horretan bertan gordeko da eta, horretaz aparte, lan eguna bukatu aurretik ikaslearen USBan gordeko da segurtasun-kopia modura.

ID	Arriskua	Gertatzeko arriskua	Eragina
A1	Kodearen galera	ertaina	handia
A2	Konputagailua hondatu	txikia	handia
A3	Planifikazioa ez betetzea	handia	ertaina

2.3 Taula: Arriskuak.

- **A2:** Ezer ez gertatzeko konputagailu pertsonala kontu handiz erabiliko da. Zorro batekin babestuta eramango da beti. Hala ere, ikaslea arrisku honen aurrean aurkitzen bada, gradu-amaierako proiektuko gelako konputagailuak eskuragarri izango ditu eta, horretaz aparte, bere etxeko konputagailua ere erabili ahal izango du.
- **A3:** Arrisku handia dago hemen, izan ere, edozein ustekabe gerta daiteke eta, horren eraginez, planifikazioa ez bete. Ustekabeak teknikoak izan daitezke (programatzerakoan erroreak, etab.) edo pertsonalagoak (ikaslea gaixorik jartzea, arazo pertsonalak, etab.). Planifikazioa ahalik eta osatuen egiten saiatuko da ikaslea hori gerta ez dadin, baina gertatzen den kasuan, emangarriak entregatzeko datak atzeratu behar badira, ez da arazorik egongo.

2.7 Komunikazio-plana

Atal honetan proiektuan zehar erabiliko den komunikazio-plana agertzen da. Lehendabizi, 2.4 taulan interesatuak zeintzuk diren azaltzen da, eta, ondoren, plana zein den.

2.7.1 Interesatuak

Pertsona	Rola	Interesa	Ezagutza maila
Maialen Otamendi	Ikaslea	Oso handia	Oso altua
Xabier Artola	Zuzendaria	Handia	Altua
Aitor Soroa	Zuzendaria	Handia	Altua

2.4 Taula: Interesatuak.

2.7.2 Komunikazio-plana

Martxan jarriko den komunikazio-planean bi modu agertzen dira. Zuzendarien eta ikaslearen arteko bilerak egingo dira, zuzendariak ikaslearen uneko egoeraren berri izateko. Bestalde, bilera egin ezin den kasuan edo biltzea merezi ez duen kasuan posta elektronikoa erabiliko da.

Bileren kasuan, zuzendariak astero bat egiteko prest daude. Hala ere, ikasleak posta elektronikoaren bidez baieztatu beharko du egitea nahi duenean. Proiektuaren erritmoa, beraz, oso-osorik ikasleak eramango du.

3. KAPITULUA

Aurrekariak

Proiektu gehienetan gertatzen den moduan, hemen egindakoa ezin zitekeen planteatu aurrekoek egin duten lana gabe. Kapitulu honetan testuinguruan murgilduko gara eta aurrekoek egin duten lana azalduko dugu, ondoren, proiektu honetan egin behar dena hobeto ulertzeko.

3.1 Testuingurua

Ixa taldeak hizkuntzaren prozesamenduaren alorrean izugarrizko lana egin du eta egiten ari da. Atal honetan talde honek erabiltzen duen anotazio-amaraunari buruzko informazioa emango da eta baita analisi linguistikoei buruzkoa ere.

3.1.1 Anotazio-amarauna

Anotazio-amarauna Ixa taldeak analisi linguistikoak errepresentatzeko erabiltzen duen datu-eredua da. XML teknologian oinarritzen da eta hizkuntzaren prozesamendua egiten duten moduluek sortutako informazioa erakutsi, irakurri eta gordetzeko erabiltzen dute.

Amaraunaren datu-ereduak eskema bat jarraitzen du. Hasierako testua eta horri lotutako informazio linguistikoa bereizita daude jarraitutako eskemaren arabera. Bien arteko erlazioa beste egitura batean gordetzen da. Hasierako testua ainguraz osatuta dago. Testuko

elementuei lotuta egongo dira aingurak, eta horietako bakoitzak bere informazio linguistikoa duen egitura erreferentziatuko du esteken bidez.

Gauzak horrela, jarraibide horien ondoren anotazio linguistikoak lortzen dira. Analisi linguistikoak elkarri loturik dauden XML dokumentuen amaraunetan adierazten dira. Amaraun horiek hiru dokumentu mota ezberdinez osaturik daude:

- **Aingurak:** Hauek corpusetako hitz fisikoak dira.
- **Anotazio-estekak:** Estekek aingura bat berari dagokion informazio linguistikoarekin lotzen dute.
- **Informazio linguistikoa:** Hitz bakoitzaren analisi linguistikoa gordetzen du.

3.1.2 Analisi linguistikoak

Analisi linguistikoak datu-egitura ezberdinak erabiliz sortzen dira. Datu-egitura horiei *Feature Structure* (FS) ere deitzen zaie. Beraz, FS bakoitzak analisi linguistiko bat errepresentatzen duela esan daiteke. Besteak beste, lema, forma, kategoria eta azpikategoria ager daitezke.

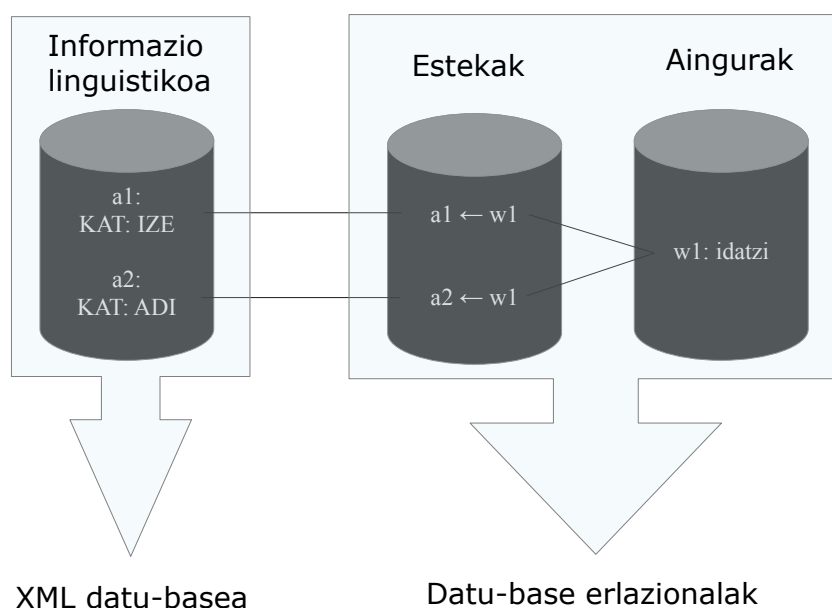
Ixa taldeak analisi linguistiko hori definitzeko edo, beste modu batera esanda, XML sintaxia jarraitzen duela eta eskema baten zehaztapenak jarraitzen dituela ziurtatzeko, RelaxNG (REGular LAnguage for XML Next Generation) eskemak erabiltzen ditu.

Aurrez esan bezala, informazioa hiru fitxategi ezberdinetan gordetzen da: analisi linguistikoak, aingurak eta bien arteko lotura ahalbidetzen duten estekak. Testu multzo handiak kontsultatu nahi diren kasuan, ordea, hau ez da bideragarria. Hori dela-eta, analisi linguistikoak XML datu-base batean gordetzea erabaki zuten, eta aingurak eta estekak, berriz, datu-base erlazional batean.

[3.1](#) irudian argiago ikusten da aurrekoa.

3.2 Armiarma

Proiektu hau Jagoba Gascón Sánchez ikasle ohiak garatu zuen Armiarma tresnan oinarritzen da, eta horren eraginkortasuna hobetzeko saiakuntza moduan aurkeztu zen. Hortaz,



3.1 Irudia: Hitzen eta analisisien arteko erlazioa datu-baseetan.

gure proiektuaren espezifikazioak azaldu aurretik, egokiena Armiarma tresnaren ezaugarriak azaltzea da.

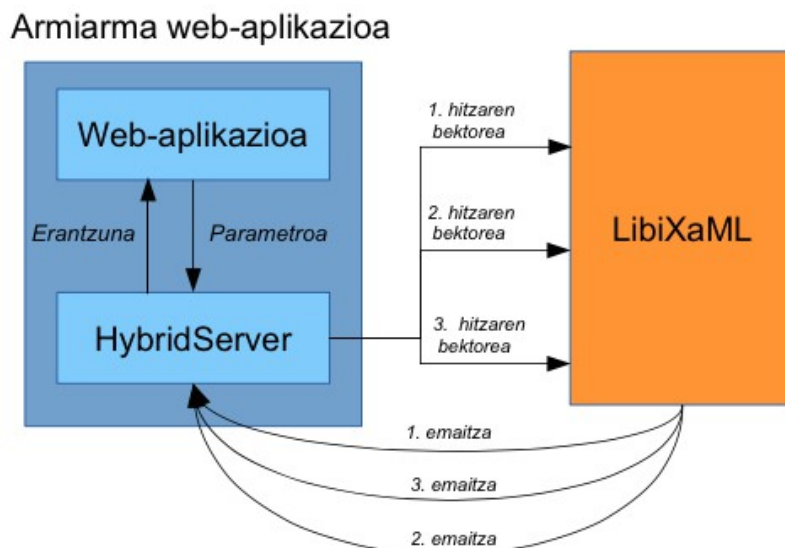
Armiarma kontsulta-tresna anotazio linguistikoak kudeatzeko web-aplikazioa da, webgunea eta C++ lengoiaz idatzitako modulu batek osatzen dutena. Tresna honek ondorengo funtzionalitateak eskaintzen dizkie erabiltzaileei: edozein motatako bilaketak egin, jasotako emaitzen artean bilatutako hitzaren analisia(k) ikusi eta analisi berriak sortu.

Bilaketa batek lau zati ditu. Lehenengoak bi parametro ditu: bat morfosintaxi-mailan lan egiteko, eta bestea analisi linguistikoak biltzeko. Beste hiruretan bilaketa-eremuak definitzen dira. Lehenengo, kontsultatu nahi den hitza definitzen da, eta, ondoren, horren testuingurua.

3.2.1 Arkitektura

Web-aplikazioa *front end/back end* motakoa da. Lehenengoa erabiltzailearen eta aplikazioaren arteko elkarrekintza ahalbidetzen duen zatia da, hau da, interfazea. Bigarrena, berriz, C++ lengoiaz idatzitako *HybridServer* moduluak eta *LibiXaML* liburutegiak osatzen dute.

3.2 irudiak arkitektura modu grafikoan azaltzen du.



3.2 Irudia: Armiaarmaren arkitektura. Iturria: Jagoba Gascón Sánchezen GAPa, 2013.

HybridServer modulua hiru zatitan banatzen da: hitzen bilaketa, analisisien eskaera eta desanbiguazioa. Aplikazioak jasotzen duen parametroa, lehendabizi, *HybridServer*era bidaltzen da. Hitz bakoitza egitura batean gordetzen da eta *LibiXaML*ra pasatzen da. Azken hau anotazio-amarauna osatzen duten fitxategi guztiak maneiatzeko liburutegia da. Datu-basean bilaketak egiten direnean, hiru emaitzak *HybridServer*era bueltatzen dira. Bertan, hitz nagusiaren erantzuna beste biekir iragazten da eta, azkenik, erantzuna aplikaziora bidali. Pantailan erabiltzaileak bilatu duen hitza agertuko da, eskatutako baldintza guztiekin eta testuinguruarekin batera.

3.3 Proiektu honi buruz

Armiaarma web-aplikazioa oso osaturik dago eta funtzionalitate asko eskaintzen dizkio erabiltzaileari. Gure proiektuan ez gara funtzionalitate gehiago gehitzen hasiko, hori ez baita helburua. Bilaketak egitera eta analisiak erakustera mugatuko gara, eta, horrez gain, aplikazioaren erantzun-denbora hobetzean murgilduko gara, izan ere, Armiaarmak duen arazo nagusia hori baita. Corpus oso handietan kontsultak egiten denbora gehiegi pasatzen denez, *Elasticsearch* tresna erabiliz erantzun-denbora hori murrizten saiatuko gara.

3.3.1 Elasticsearch

Informazioaren berreskurapenerako erabiltzen den motorra da, dokumentuei orientatutakoa, eta *Apache Lucenen* oinarritzen da. Datu-multzo handiak indexatu eta analizatzeko aukera ematen du denbora gutxian.

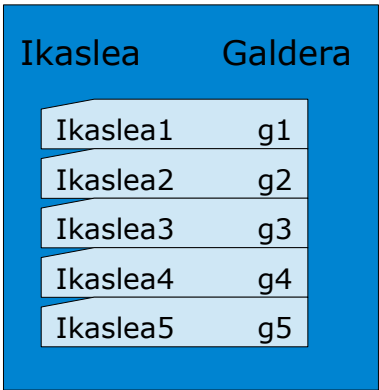
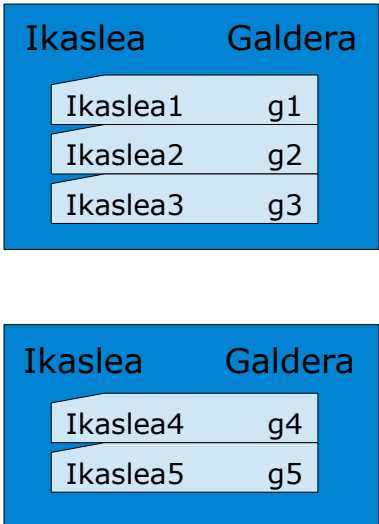
Bi ezaugarri nagusi ditu:

- **Sistema banatua:** Informazioa makina eta multzo ezberdinetan banatzeko pentsatuta sortu zen.
- **Bilaketak egiteko motorra:** Indexazioak eta bilaketak egiteko funtzionalitateak eskaintzen ditu.

Gehiago sakontzen hasi aurretik, kontzeptu batzuk definituko dira, ondoren datorren guztia ulergarriagoa izan dadin.

- **Cluster:** *Elasticsearchen* instantzia multzoa da, edo, beste modu batera esanda, nodo multzoa.
- **Node:** *Elasticsearchen* instantzia.
- **Index:** Dokumentu ezberdinen bilduma da, datu-base erlazioaletan datu-base kontzeptuaren parekoa. Dokumentuek JSON formatukoak izan behar dute, baina ez da beharrezkoa denek egitura bera izatea.
- **Type:** Egitura bera duten dokumentuen bilduma da, datu-base erlazioaletan taula moduan ezagutzen dena. *Index* batek *type* bat edo gehiago izan ditzake. Horregatik azpimarratu da aurreko puntuan ez dela beharrezkoa *index* bateko dokumentuek egitura bera izatea, *type* bat edo gehiago egon daitezkeelako (bakoitzean).
- **Shard:** Informazioa hainbat zatitan bana daiteke, edo, beste modu batera esanda, dokumentuak multzo ezberdinetan bana daitezke. Multzo bakoitzari *shard* deitzen zaio. *Node* batek *shard* bat edo gehiago administra ditzake. 3.3 irudia ikus daiteke hau ondo ulertzeko. Adibidearen ezker aldean *shard* bat agertzen da bost dokumenturekin (dokumentu bakoitzak ikaslea eta galdera eremuak ditu) eta eskuinaldean, berriz, dokumentu horiek bi *shard*etan banatuta daudela ikus daiteke.
- **Replica:** Informazio zati bakoitzaren (*shard*) kopia.

- **Mapping:** *Type* batek izango duen datu-egitura zehazten du. *Index* batek *type* bat baino gehiago baldin baditu, bakoitzaren datu-egitura definitu ahal izango da. Hor-taz, gerta daiteke indize batean *mapping* bat baino gehiago definitzea.

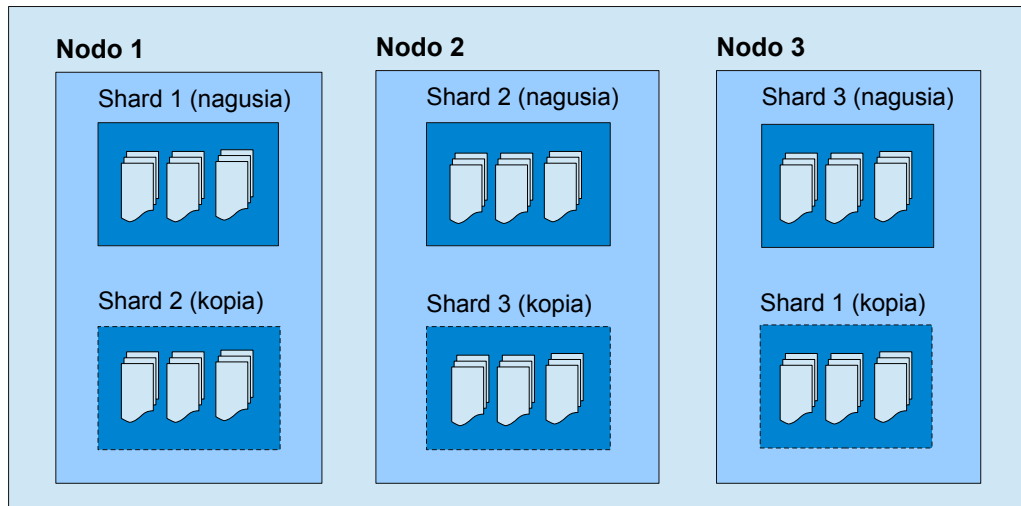
<i>Shard</i> bat	Bi <i>shard</i>																										
 <table border="1"> <thead> <tr> <th>Ikaslea</th><th>Galdera</th></tr> </thead> <tbody> <tr><td>Ikaslea1</td><td>g1</td></tr> <tr><td>Ikaslea2</td><td>g2</td></tr> <tr><td>Ikaslea3</td><td>g3</td></tr> <tr><td>Ikaslea4</td><td>g4</td></tr> <tr><td>Ikaslea5</td><td>g5</td></tr> </tbody> </table>	Ikaslea	Galdera	Ikaslea1	g1	Ikaslea2	g2	Ikaslea3	g3	Ikaslea4	g4	Ikaslea5	g5	 <table border="1"> <thead> <tr> <th>Ikaslea</th><th>Galdera</th></tr> </thead> <tbody> <tr><td>Ikaslea1</td><td>g1</td></tr> <tr><td>Ikaslea2</td><td>g2</td></tr> <tr><td>Ikaslea3</td><td>g3</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Ikaslea</th><th>Galdera</th></tr> </thead> <tbody> <tr><td>Ikaslea4</td><td>g4</td></tr> <tr><td>Ikaslea5</td><td>g5</td></tr> </tbody> </table>	Ikaslea	Galdera	Ikaslea1	g1	Ikaslea2	g2	Ikaslea3	g3	Ikaslea	Galdera	Ikaslea4	g4	Ikaslea5	g5
Ikaslea	Galdera																										
Ikaslea1	g1																										
Ikaslea2	g2																										
Ikaslea3	g3																										
Ikaslea4	g4																										
Ikaslea5	g5																										
Ikaslea	Galdera																										
Ikaslea1	g1																										
Ikaslea2	g2																										
Ikaslea3	g3																										
Ikaslea	Galdera																										
Ikaslea4	g4																										
Ikaslea5	g5																										

3.3 Irudia: *Shard* bat eta bi *Sharden* arteko diferentzia.

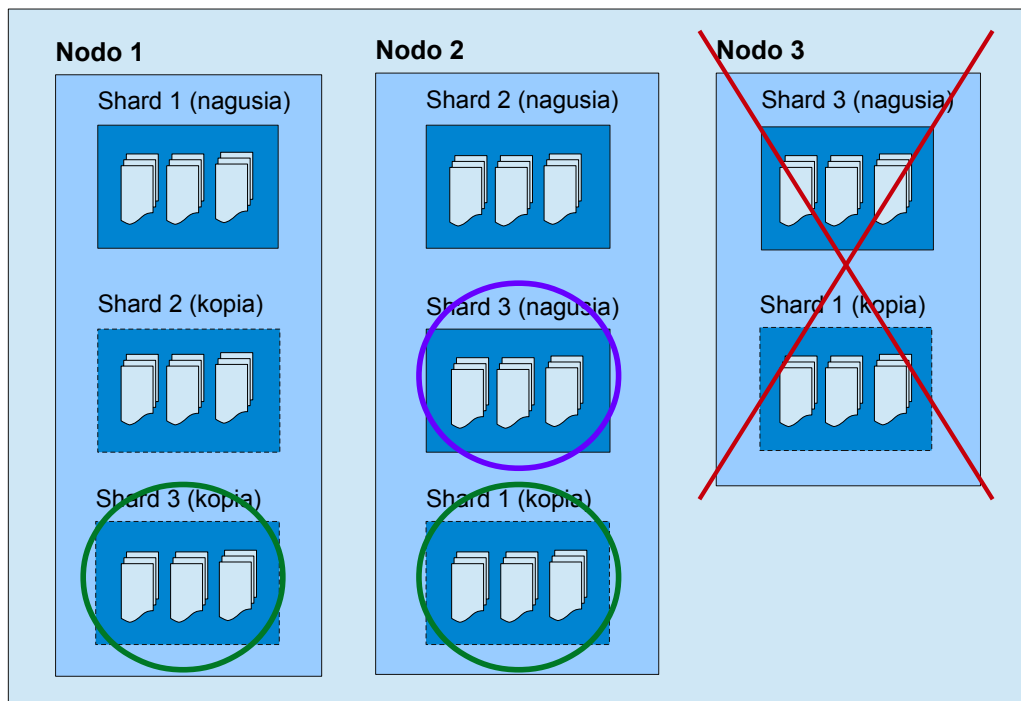
Hasiera-hasieratik, *Elasticsearch*ek *cluster* bat sortzen du. Ondoren, *index* berri bat sortzean zenbat *shard* eta *replica* nahi diren zehatz daiteke. Hori zehazten ez den kasuan, besterik ezean 5 *shard* eta *replica* bat jarriko dira.

Kontzeptu horiek guztiak hobeto ulertzeko begiratu 3.4 irudia. Adibidean agertzen dena *cluster* baten egitura ohiko bat da. Hiru nodoz, hiru *shardez* eta horietako bakoitzaren kopia batez (*replica*) osaturik dago. Azpimarratzekoa da *shard* baten kopia ez dela inoiz *shard* nagusia dagoen tokian gordetzen. *Sharden* kopiak nodoren bat erortzen den kasuan bertako informazioa ez galtzeko daude, eta, horretarako, ezinbestekoa da kopiak beste nodoetan gordetzea.

3.5 irudian nodoren bat erortzen bada zer gertatzen den agertzen da. Hirugarren nodoak hirugarren *shard* nagusia eta lehenengo *sharden* kopia zituen gordeta. Nodo hori erortzen

CLUSTER

3.4 Irudia: *Elasticsearch*en kontzeptuak ulertzeko irudia.

CLUSTER

3.5 Irudia: Nodoren bat erortzen bada zer gertatzen den erakusten duen irudia.

bada, *Elasticsearch*ek hirugarren *shard* nagusia falta dela sumatzen duenean, bere kopia *shard* nagusi batean bihurtzen du (biribil morez agertzen dena). Ondoren, lehenengo eta hirugarren *sharden* kopiak falta direnez, bakoitza nodo batean gordetzen ditu (biribil berdez agertzen dena). Horrela, dokumentuetan gordetzen den informazioa ez galtzea lortzen da.

Bezeroak eskaera bat egiten duenean edozein nodori egin diezaioke, nodo guztiak baitira gai edozein eskaerari erantzuteko.

3.3.2 Datu-base erlazionalekiko desberdintasunak

Orain arte gradu osoan zehar egin ditugun lan gehienak datu-base erlazionalekin egin dira; beraz, aurreko hitzen definizioak hobeto uler daitezten, 3.1 taulan datu-base erlazionalekiko ezberdintasunak modu garbian agertzen dira, kontzeptuen arteko baliokidetzak zerrendatuz.

Datu-base erlazionala	Elasticsearch
Datu-basea	Index
Taula	Type
Lerroa	Document
Zutabea	Field
Taulen definizioa	Mapping

3.1 Taula: Datu-base erlazionalekiko desberdintasunak.

3.3.3 Funtzionamendua

Elasticsearch martxan jartzeko, instalazioa burutu ondoren, hauek dira jarraitu beharreko pausoak:

1. **Index berri bat sortu.** Hori datu-egitura edo *mappinga* zehaztuz egiten da.
2. **Datuak sartu.** Horretarako, aurrerago azalduko den *Index* APIa erabiltzen da.
3. **Bilaketak egiten hasi.** Horretarako, aurrerago azalduko den *Get* APIa erabiltzen da

Esan beharra dago zerrendatu diren hiru pauso horietatik lehenengoa ez dela beharrezkoa egitea. Zuzenean datuak indexatzen badira, *Elasticsearch*ek berak ulertzen duen eskema

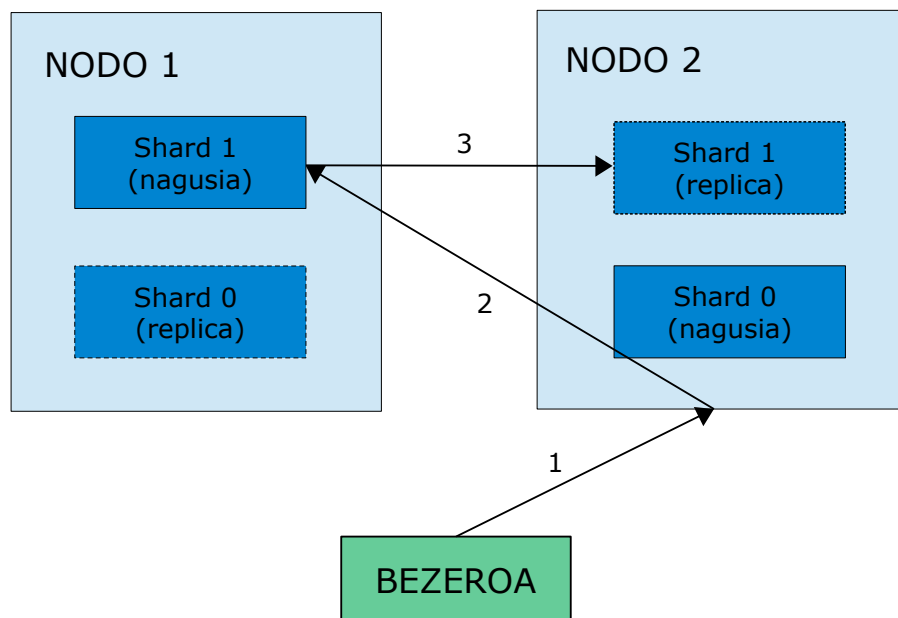
sortzen du, datu horiek jarraitzen duten egituraren arabera. Hala ere, komeni da *mappinga* definitzea datu-egitura zehatz bat izatea nahi bada, behintzat. Gure kasuan horrela izan da.

Jarraian, datuak nola indexatzen diren eta bilaketak nola egiten diren azalduko da, *Index* eta *Get* APIak erabiliz.

Index APIa

Index APIak JSON dokumentu bat gehitu edo editatzen du indize batean. 3.6 irudian ikus daitezke dokumentu bat gehitzean ematen diren pausoak.

1. Nodo jakin bati eskaera egiten zaio (dokumentua indexatu nahi duela).
2. Zerbitzariak eskaera irakurtzen du, eta dokumentua zein *shard* nagusitan gordeko duen erabakitzen du.
3. Dokumentua *shard* nagusian gorde ondoren, kopiak egiten dira (*replicak*).

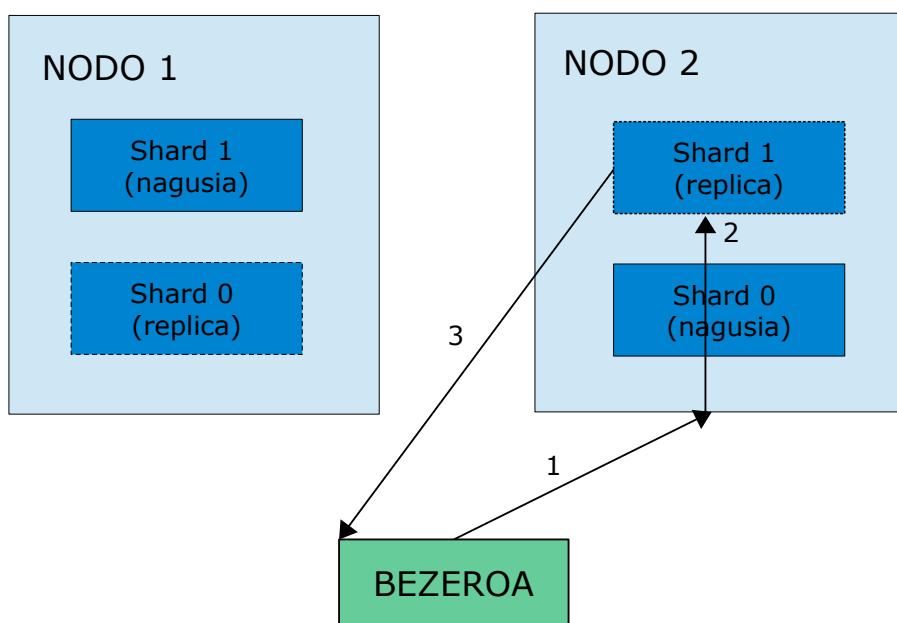


3.6 Irudia: *Index* APIaren funtzionamendua

Get APIa

Get APIak gordeta dagoen dokumentu bat lortzen du eta, horretarako, dokumentuaren identifikadorean oinarritzen da. 3.7 irudian ikus daitezke dokumentu bat jasotzean ematen diren pausoak.

1. Bezeroak GET eskaera bat bidaltzen dio nodoari.
2. Nodoak eskaera irakurtzen du, eta dokumentua zein *shard*etan gordeta dagoen erabakitzen du.
3. Dokumentua jasotzen da.



3.7 Irudia: *Get* APIaren funtzionamendua

4. KAPITULUA

Diseinua

Proiektuaren diseinua egiterakoan bi atal ezberdinetan zentratu behar izan gara. Lehena, eta garrantzia handiena duena, datu-egituraren diseinua. Bigarrena, web-aplikazioaren interfazea. Nolabaiteko oreka lortu behar zen interfazearen funtzionalitateen eta eraginkortasunaren artean. Horregatik, diseinua egitea ez da lan erraza izan.

4.1 Elasticsearch

Hasieratik garbi geneukana zera zen: hitz bat bilatzerakoan bere analisia, aurreko eta hurrengo bost hitzen analisiak, eta aurreko eta ondoko esaldiak atzitzeko gai izan behar zuela aplikazioak. Hortaz, hori guztia ahalbidetzen duen datu-egitura bat behar zen.

4.1.1 Dokumentuen eta analisisien diseinua

Hitzen analisisiek ezaugarri asko dituzte eta zein formatutan agertuko diren ere pentsatu beharreko zerbait zen. Analisi baten ezaugarri nagusienak hiru dira: forma, lema eta kategoria. Kategoriaren kasuan hamabi ezberdin daude, eta bakoitzak bere ezaugarri propioak ditu. Armiarma aplikazioarekin alderatuz ezaugarri gutxiago sartu dizkiogu, aurrez esan bezala, ez baitzen hori gure helburua. Hona hemen kategoriak eta bakoitzaren ezaugarriak:

- **Izena:** Azpikategoria, kasua, mugatasuna eta numeroa.

- **Aditza:** Azpikategoria, aditz mota, aditzoina, aspektua, erlazioa, kasua, laguntzaile mota, modua-denbora eta numeroa.
- **Adjektiboa:** Azpikategoria, izenaren aurretik, kasua, mugatasuna eta numeroa.
- **Adberbioa:** Azpikategoria, adberbio mota, kasua, gradu-maila, mugatasuna eta numeroa.
- **Izenordaina:** Azpikategoria, kasua, mugatasuna, numeroa-mugatasuna, numeroa eta pertsona.
- **Aditz laguntzailea:** Erlazioa, hitanoa, kasua, modua-denbora, mugatasuna, nor, nori eta nork.
- **Aditz trinkoa:** Aspektua, erlazioa, hitanoa, kasua, modua-denbora, mugatasuna, nor, nori eta nork.
- **Determinatzailea:** Azpikategoria, kasua, mugatasuna, numeroa-mugatasuna eta numeroa.
- **Interjekzioa:** Azpikategoria, kasua eta mugatasuna.
- **Loturazkoa:** Azpikategoria, erlazioa eta klausula-muga.
- **Partikula:** Erlazioa eta modalitatea.
- **Bestelakoa:** Azpikategoria, kasua eta mugatasuna.

Behin ezaugarri guztiak zehaztuta, horiek osatuko zuten analisiaren formatua zehaztu behar zen. Hainbat proba egin genituen, baina azkenean Ixa Taldekoek diseinatutako formatua erabiltzea erabaki zen. Beraiek zehaztuta daukaten moduan, analisia bi marra etzanez ("/") banatutako ezaugarri linguistikoen multzoa da, eta ezaugarri bakoitzak bere posizio zehatza du. Ezingo da ezaugarri baten posizioa aldatu, dokumentuetako datu guztiak formatu horretan baitaude.

//Forma//Lema//Kategoria//Mugatasuna//Kasua//Azpikategoria//Numeroa//Gradu-maila//Kasumetatzea//Erlazioa//Modalitatea//Aditzoina//Nork//Nori//Nor//Modua-denbora//Aspektua//Numeroa-mugatasuna//Metakategoria//Laguntzaile mota//Izenaren aurretik//Hitanoa//Aditz mota//Plurala//Pertsona//Klausula-muga//Adberbio mota//Analisi zuzena edo ez

Aurreko adibideko guztia izango litzateke hitz baten analisi bat, baina, noski, hitz batek ez ditu ezaugarri guztiak izango. Hortaz, hitz bati horietako ezaugarriren bat ez badagokio, hutsa uzten da bere lekua (4.1 irudian urdinez markatuta agertzen da analisi baten adibidea).

Analisei buruz azaldu beharreko guztia azalduta, 4.1 irudian ikus daiteke dokumentu batean hitzak nola agertzen diren.

```
analisia://hau//hau//DET//M//ABS//ERKARR//S//////////o_ez
//hau//*edun//ADL////////HARK//HI/A1//////////o_ez
//hau//ukan//ADT////////HARK//HI/A1/PNT//////////o_ez
aurrekoak1=//herrialde//herrialde//IZE////ARR//////////o_z
//herrialde//herrialde//IZE/MG/ABS//ARR//////////o_ez
aurrekoak2=//dituen//ukan//ADT////////MOS////HARK//HAIEK//A1/PNT//////////o_z
//dituen//*edun//ADL////////ERLT////HARK//HAIEK//A1//////////o_ez
//dituen//*edun//ADL////////ZHGH////HARK//HAIEK//A1//////////o_ez
//dituen//*edun//ADL////////MOS////HARK//HAIEK//A1//////////o_ez
//dituen//ukan//ADT////////ERLT////HARK//HAIEK//A1/PNT//////////o_ez
//dituen//ukan//ADT////////ZHGH////HARK//HAIEK//A1/PNT//////////o_ez
aurrekoak3=//aintzira//aintzira//IZE//M//ABS//ARR//S//////////o_z
//aintzira//aintziratu//ADI/////SIN////////aintzira////////DU////////ADOIN//////////o_ez
//aintzira//aintzira//IZE////ARR//////////o_ez
//aintzira//aintzira//IZE/MG/ABS//ARR//////////o_ez
aurrekoak4=//188.000//188.000//DET//MG/ABS//DZHH////////P//////////o_z
//188.000//188.000//DET//MG/ABS//DZHH////////P//////////o_ez
//188.000//188.000//IZE////ZKI//////////o_ez
//188.000//188.000//IZE/MG/ABS//ZKI/S//////////o_ez
//188.000//188.000//IZE//M//ABS//ZKI/S//////////o_ez
aurrekoak5=//eta//eta//LOT////////JNT////////EMEN//////////o_z
//eta//eta//LOT////////MEN////////KAUS////////AM//////////o_ez
hurrrangoak1=//errepublika//errepublika//IZE////ARR//////////o_z
//errepublika//errepublika//IZE/MG/ABS//ARR//////////o_ez
//errepublika//errepublika//IZE/M//ABS//ARR//S//////////o_ez
hurrrangoak2=//independentea//independente//ADJ//M//ABS//ARR//S//////////o_z
hurrrangoak3=//da//izan//ADT////////HURA//A1/PNT//////////o_z
//da//izan//ADL////////HURA//A1//////////o_ez
hurrrangoak4=//1917az//1917//ZE//M//INS//ZKI/S//////////o_z
hurrrangoak5=//geroztik//geroztik//ADB////////ARR//////////o_z
aurrekoa=Konifero baso erraldoiak eta 188.000 aintzira dituen herrialde
ondokoa= ,errepublika independentea da 1917az geroztik. Suedia eta ...
```

4.1 Irudia: Dokumentu bateko hitz baten adibidea.

Hitz bakoitzak hamahiru eremu ditu: analisia, aurrekoak1, aurrekoak2, aurrekoak3, aurrekoak4, aurrekoak5, hurrengoak1, hurrengoak2, hurrengoak3, hurrengoak4, hurrengoak5, aurrekoa eta ondokoa. Lehenengo hamaika eremuek hitz desberdin baten analisiak errepresentatzen dituzte. Analisia eremuak bloke horren hitzaren (hitz nagusia) analisia jasotzen du. AurrekoakX (X 1etik 5erako edozein zenbaki izanik) eremuek hitz nagusia baino X posizio atzerago dagoen hitzaren analisiak jasotzen dituzte. HurrengoakX (X 1etik 5erako edozein zenbaki izanik) eremuek hitz nagusia baino X posizio aurrerago dagoen hitzaren analisiak jasotzen dituzte. Bukatzeko, aurrekoa eta ondokoa eremuek hitz nagusiaren aurretik eta ondoren dauden esaldien zati bat jasotzen dute.

Aipatzekoa da analisiaren eremu bakoitza balio zehatz batekin espezifikatzen dela. Balio horiek eranskinean dagoen fitxategian ikus daitezke nahi izanez gero.

4.1.2 *Shard* eta *replica* kopuruak zehaztu

*Elasticsearch*i buruzko azalpena egin denean zati bat bi kontzeptu hauei eskaini diegu, garrantzia handia du eta. Hori ondo ulertu ondoren, gure proiektuak bakoitzetik zenbat izango zituen pentsatu behar zen.

- **Shards:** Eskaera oso bat prozesatzeko *shard* guztietan prozesatu behar da. Nodo bakoitzeko *shard* bakarra izanez gero, eskaera guztiak paraleloan exekuta daitezke. Nodo batean *shard* bat baino gehiago izanez gero, berriz, eskaerak seriean exekuta beharko lirатеke. Hori jakinda, *shard* bakarra erabiltzea erabaki genuen.
- **Replica:** Aktibo dauden *sharden* kopiak izaten dira eta nodo batek erantzuten ez duen kasurako egokia da gutxienez *shard* bakoitzeko kopia bat gordetzea. Horregatik, kopia bakarra esleitzea erabaki genuen.

```
"settings" : {  
  "number_of_shards" : 1,  
  "number_of_replicas" : 1  
}
```

4.2 Irudia: Guk definitutako *shard* eta *replica* kopuruak.

4.1.3 Datu-egituraren diseinua

Datu-egitura *mappinga* definituz zehazten da. *Mappinga Elasticsearch*en kontzepturik garrantzitsuenetakoa da, bilaketa-motorrak dokumentuen prozesamendua nola egin behar duen zehazten baitu.

Indize berri bat sortzerakoan *mappinga* definitu ez bada, defektuz sortzen da dokumentu bat osatzen duten datuen egitura jarraituz. Ondoren, *mapping* berria nodoetan zehar banatzen da. Gure kasuan aurrez definitu dugu, nahiko datu-egitura zehatza nahi genuelako.

4.3 irudian guk definitutako *mapping* edo datu-egitura ikus daiteke.

```

"mappings" : {
  "_default_" : {
    "properties" : {
      "analisia" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoak1" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoak2" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoak3" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoak4" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoak5" : {
        "type": "string", "analyzer": "whitespace" },
      "hurrrangoak1" : {
        "type": "string", "analyzer": "whitespace" },
      "hurrrangoak2" : {
        "type": "string", "analyzer": "whitespace" },
      "hurrrangoak3" : {
        "type": "string", "analyzer": "whitespace" },
      "hurrrangoak4" : {
        "type": "string", "analyzer": "whitespace" },
      "hurrrangoak5" : {
        "type": "string", "analyzer": "whitespace" },
      "aurrekoa" : {
        "type": "string", "index": "not_analyzed" },
      "ondokoa" : {
        "type": "string", "index": "not_analyzed" }
    }
  }
}

```

4.3 Irudia: Guk definitutako datu-egitura edo *mapping*a.

- **analisia, aurrekoak eta hurrrangoak:** Ereku hauek guztiek hitzen analisia gordetzen dituzte: *analisia* eremuak hitz nagusiaren analisia gordetzen du. *Aurrekoak1*, *aurrekoak2*, *aurrekoak3*, *aurrekoak4* eta *aurrekoak5* eremuek, hitz horren aurretik dauden analisia gordetzen dituzte eta *hurrrangoak1*, *hurrrangoak2*, *hurrrangoak3*, *hurrrangoak4* eta *hurrrangoak5* eremuek, berriz, hitz nagusiaren ondoren dauden analisia.
- **type:** Ereku honek *string* motakoa izan behar du.
- **analyzer:** Kontzeptu hau *Lucenetik* dator. Testua tokenetan banatzen duen tokenizatzaile batekin osatzen da. Gure kasuan, eremu horietako bakoitzak analisi bat baino gehiago izan dezake. Hori kontuan hartuta, nolabait desberdindu behar zen analisi bakoitza. Hortaz, *whitespace* analizatzailea erabiltzea erabaki genuen, eremu bakoitzeko *string* osoa (analisi guztiak) zuriunez banatu ahal izateko.
- **Aurrekoa eta ondokoa:** Web-aplikazioaren interfazeaz, bilatutako hitz nagusia testuinguruan agertu behar denez, *aurrekoa* eta *ondokoa* eremuek hitz nagusiaren bi alboetan dauden esaldiak gordetzen dituzte. Hala ere, ez dira esaldi osoak izango (gutxi gora behera bakoitzean hamar hitz gordetzen dira).
- **type:** Ereku honek ere *string* motakoa izan behar du.

- **index:** Kasu hauetan ez genuen testua zuriunetan banatzea nahi, esaldi oso baten moduan interpretatzea interesatzen zitzaigulako. Hortaz, aurrekoa eta ondokoa eremuak ez zutela inongo analizatzailek izango erabaki genuen.

4.1.4 Bilaketak

Aplikazioak hiru hitz bilatzeko aukera eman behar du eta beraien arteko distantzia erabakitzeko aukera ere bai. Hortaz, datu-egitura zehaztu ondoren, *Elasticsearch*ek horretarako ematen dituen aukeren artean bilaketak egiteko modua zehaztu zen.

Kontsulta boolearrak erabili dira bilaketak egiteko modua diseinatzeko. Mota honetakoak dira gehien erabiltzen direnak, izan ere, eskaera oso handiak diseinatzeko aukera ematen dute eta.

Bestalde, *wildcard* espresioak erabili dira analisisien inguruko bilaketak egiteko, baina kontu hori ez da bat ere zaila izan, analisisien formatuak ez baitu zailtasunik. Horri erreparatzen badiogu, bi marra etzanez ("/") daude banatuta ezaugarri linguistikoak eta ezaugarri bakoitzak bere posizio zehatza du. Beraz, espresioa osatzeko, bilatu nahi den hitzaren ezaugarri bakoitza bere posizioan jarri beharko da, eta hutsik geratzen diren zatietan *edotzer* egon daitekeela adierazi behar da. Azken hori, "*" karakterearekin zehazten da. Ikur horrek edozein karaktere-sekuentzia onartzen du.

Esan beharra dago espresio erregularrak ere erabil daitezkeela, baina horiek erabilita aplikazioaren erantzun-denbora handiagoa dela. Gainera, ez ditugu behar espresio horiek ematen dituzten funtzionalitate guztiak.

Hiru hitz bilatzeko aukera eskaini behar genuenez, hiru kontsulta mota desberdindu genituen. Bakoitza adibide batekin azalduko dugu jarraian:

- **Hitz batekin** (4.4 irudia). Hitz bateko bilaketek ez dute misterio handirik. `must` eremuaren barruan dagoen *wildcard* espresioa derrigorrez bete beharko da. Adibidean `forma='Baina'` duten hitzak eskatzen dira.
- **Bi hitzekin** (4.5 irudia). Hemen bi zati nabarmentzen dira `must` eremuaren barruan. Bata lehenengo hitzari dagokio eta bestea bigarrenari. Lehenengoak bere barnean beste `must` bat du, izan ere, lehenengo hitza beti aurkitu behar baita. Bigarren hitzari dagokionez, erabiltzaileak nahi duen distantzia (1etik 5era) eta kokapenaren (aurretik, atzetik edo edonon) arabera (adibideko kasuan, lehenengo hitzetik 1eko

distantziara eta edonon egotea nahi zen), *wildcard* espresio bat edo gehiago egongo dira. Horietako bat gutxienez bete beharko da emaitza itzultzeko. Adibidean, lehenengo hitzaren forma *Baina* izatea nahi da eta, horretaz aparte, bere aurreko edo atzeko hitzaren forma *ez* izatea.

- **Hiru hitzekin** (4.6 irudia). Hemen hiru zati nabarmentzen dira *must* eremuaren barruan. Bata lehenengo hitzari dagokio, bestea bigarrenari eta azkena hirugarrenari. Lehenengoa, betiko moduan, beti bete beharko da, eta horregatik *doa must* eremuaren barruan. Beste biak *should* eremuaren barruan *doaz*, eta horrek esan nahi du horietatik gutxienez bakarra bete beharko dela. Adibidean, lehenengo hitzaren forma *Baina* izatea nahi da. Bigarren hitzari dagokionez (forma *ez duena*), lehenengotik leko distantziara aurretik edo atzetik egotea nahi da. Azkenik, hirugarren hitza (forma *dezagun duena*) lehenengotik bi posizio aurrerago edo atzerago egotea nahi da. Kontuan izan behar da distantzia kopurua jartzen denean, distantzia hori edo gutxiago betetzen duten emaitzak itzuli behar direla, horrela interpretatu nahi izan dugulako distantziaren kontzeptua; horregatik agertzen dira adibide honetan *aurrekoak1* eta *hurrengoak1* eremuak ere bai).

[illegible]

4.4 Irudia: Hitz bat bilatzeko eskaeraren adibidea.

4.2 Web-aplikazioaren interfazea

3. orrialdean interfazeak bete behar zituen ezaugarriak zerrendatu ditugu eta ezaugarri horiek guztiak eskainiko zituen interfaze bat behar zen. 4.7 irudiak web-aplikazioaren interfazea hasierako egoeran erakusten du. 4.8 eta 4.9 irudietan emaitzak eta analisiak nola ikusten diren agertzen da. 4.8 irudian bilaketa baten emaitzak agertzen dira eta ikus daiteke ez direla emaitza guztiak orrialde bakarrean kargatzen. Emaitza kopuruaren arabera, orrialdeak sortzen dira eta bakoitzak hogeit hamar emaitza bakarrik erakusten ditu, erabiltzailearentzat erosoagoa izan dadin.

4.5 Irudia: Bi hitz bilatzeko eskaeraren adibidea.

4.6 Irudia: Hiru hitz bilatzeko eskaeraren adibidea.

The screenshot shows a web application interface with a dark header bar containing the text "Emitzak" and "Analisia". Below the header, there is a large gray rectangular area labeled "Emitzak:". To the right of this area is a red sidebar containing a search form. The form has a title "1 Hitzza" and a question "Analisi zuzena nahi al duzu?". It includes input fields for "Baina" (containing "Baina"), "Lema" (containing "Lema"), and a dropdown menu for "--Kategoria--". Below the input fields, there is a section titled "Bigarren hitza bilatu:" with a radio button labeled "BAI" and two buttons labeled "Bilatu" and "Garbitu".

4.7 Irudia: Web-aplikazioaren interfazea: "Baina"lemaren bilaketa.

The screenshot shows the same web application interface as before, but now displaying search results. The "Emitzak:" label is followed by the number "617". Below this, there is a large block of text containing multiple occurrences of the word "Baina" in various contexts. At the bottom of the text block, there is a pagination control showing "<< < 31-tik 2 > >>".

4.8 Irudia: Web-aplikazioaren interfazea 4.7 irudiko bilaketaren emaitzekin.

Emailtzak

Analisia

Analisi zuzena ✓

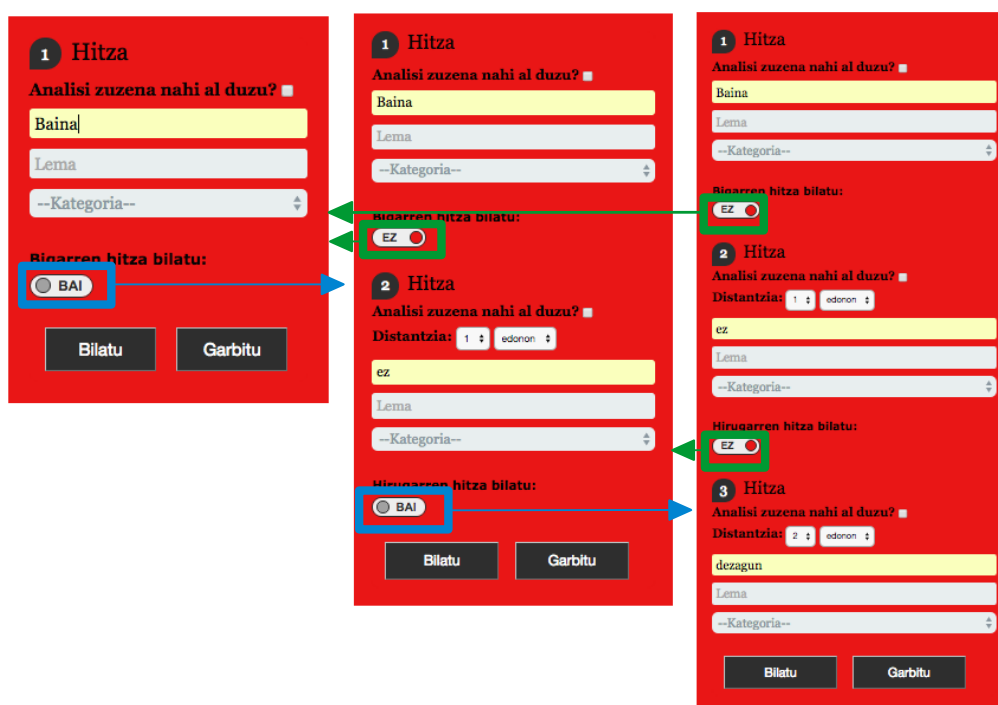
Forma: Baina
Lema: baina
Kategoria: Loturazkoa
Azpikategoria: Juntagailua
Erlazioa: Aurkaritzakoa

Analisi okerra ✗

Forma: Baina
Lema: bainatu
Kategoria: Aditza
Azpikategoria: Aditz sinplea
Aditzoina: baina
Aditz mota: Aditzoina

4.9 Irudia: Web-aplikazioaren interfazea hitz baten analisiarekin.

Formularioari erreparatzen badiogu (ikusi 4.10 irudia), ikus daiteke bigarren hitza bilatzeko aukera eskaintzen duela *BAI* botoia sakatuz, eta, horrekin batera, hirugarren hitza bilatzeko aukera ere ematen du. *BAI* jartzen duen bigarren botoian sakatuz gero, hirugarren blokea agertuko da.



4.10 Irudia: Interfazeko formularioaren funtzionamendua.

4.2.1 Formulario dinamikoa

Formularioak, neurri handi batean, dinamikoa izan behar duela esan dugu. Hitz baten kategoria aukeratzekoan, aplikazioak kanpoko fitxategi batetik kategoria horri dagozkion ezaugarriak atzitzen ditu eta interfazeerako eremuak sortzen ditu. Beraz, formularioko ezaugarriak sortzeko informazioa deklaratiiboki gordetzen da, kodean txertatu gabe.

Horretarako, zuzendariak ikasleak Armiarman erabiltzen duten *RelaxNG* fitxategi bat pasatu zioten eta ikasleak JSON formatura bihurtu behar izan zuten (eranskinen atalean dago fitxategiaren edukia). 4.11 irudian *Izena* kategoriaren ezaugarriak agertzen dira. Beste kategoria guztietarako egitura bera erabili zen.

Ezaugarri bakoitzari balioa eta izena eremuak sartu zaizkiola ikus daiteke. Bata, dokumentuetan dauden analisi guztiak balio horiekin egon behar direlako gorde behar da,

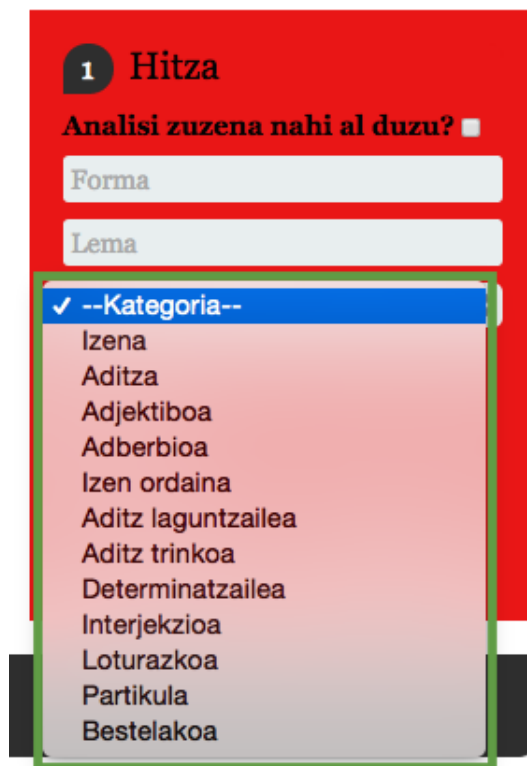
```

{
  "Kategoriak" : [
    {
      "balioa" : "IZE",
      "izena" : "Izena",
      "ezaugIzenak" : ["Azpikategoriak", "Kasua", "Mugatasuna", "Numeroa"],
      "Ezaugarriak" : [
        {
          "Azpikategoriak" : [
            {
              "balioa" : "ARR", "izena" : "Izen arrunta",
              "balioa" : "LIB", "izena" : "Leku-izen berezia",
              "balioa" : "IZB", "izena" : "Pertsona-izen berezia",
              "balioa" : "ZKI", "izena" : "Zenbakia"
            }
          ],
          "Kasua" : [
            {
              "balioa" : "ABL", "izena" : "Ablatiboa",
              "balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa",
              "balioa" : "ABS", "izena" : "Absolutiboa",
              "balioa" : "ABZ", "izena" : "Adlatibo bide zuzeneko",
              "balioa" : "ALA", "izena" : "Adlatiboa",
              "balioa" : "BNK", "izena" : "Bankaria",
              "balioa" : "DAT", "izena" : "Datiboa",
              "balioa" : "DES", "izena" : "Destinatiboa",
              "balioa" : "DESK", "izena" : "Deskribatzailea",
              "balioa" : "ERG", "izena" : "Ergatiboa",
              "balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa",
              "balioa" : "GEN", "izena" : "Genitibo edutezkoa",
              "balioa" : "INE", "izena" : "Inesiboa",
              "balioa" : "INS", "izena" : "Instrumentala",
              "balioa" : "MOT", "izena" : "Motibatiboa",
              "balioa" : "PAR", "izena" : "Partitiboa",
              "balioa" : "PRO", "izena" : "Prolatiboa",
              "balioa" : "SOZ", "izena" : "Soziatiboa"
            }
          ],
          "Mugatasuna" : [
            {
              "balioa" : "M", "izena" : "Mugatua",
              "balioa" : "MG", "izena" : "Mugagabea"
            }
          ],
          "Numeroa" : [
            {
              "balioa" : "S", "izena" : "Singularra",
              "balioa" : "P", "izena" : "Plurala",
              "balioa" : "PH", "izena" : "Plural hurbila"
            }
          ]
        }
      ]
    }
  ]
}

```

4.11 Irudia: *Izena* kategoriaren ezaugarriak.

eta, bestea, formularioan izen osoak agertu behar direlako gorde behar dira. Erabiltzaileak ez baitu zertan jakin balio bakoitza zer den. 4.12 irudian ikus daiteke nola ikusten diren ezaugarrien izenak.



1 Hitza

Analisi zuzena nahi al duzu? ☐

Forma

Lema

✓ --Kategoria--

- Izena
- Aditza
- Adjektiboa
- Adberbioa
- Izen ordaina
- Aditz laguntzailea
- Aditz trinkoa
- Determinatzailea
- Interjekzioa
- Loturazkoa
- Partikula
- Bestelakoa

4.12 Irudia: Formularioan ezaugarri lingustikoen izen osoak agertzen dira.

5. KAPITULUA

Proiektuaren garapena

Kapitulu honetan proiektuaren garapena azalduko da. Nahiz eta garapen nagusia web-aplikazioa izan, datuen formatua aldatzeko edota datuen indexazioa automatikoki egiteko scriptak ere garatu dira. Hori dena azaltzen saiatuko gara hemen.

5.1 Datuen formatua aldatu

Zuzendariek ikasleari datuak formatu batean pasatu zizkioten, baina horiek indexatzeko formatuz aldatu behar zituen. Diseinuko atalean dagoen [4.1](#) irudian ikus daiteke zuzendariek ikasleari pasatutako dokumentu baten bloke bat.

Dokumentu asko indexatu behar zirenez, hori egiteko modurik egokiena *Bulk* APIa erabiltzea zen. API hori indexatze asko dei bakar batean errepresentatzeko gai da. Indexatzeko modu honek dokumentuek egitura zehatza izatea eskatzen du, eta egitura hori [5.1](#) irudian ageri da. Esan beharra dago lerro-saltoak errespetatu egin behar direla, eta dokumentuko bloke bakoitzaren gainean datuak indexatu nahi ditugula adierazi behar da (gorriz azaldu-ta dagoen moduan, izan ere, *index* aginduaz aparte, *delete*, *create* eta *update* aginduak ere sar daitezke). Horretaz aparte, zein indizetan gorde behar den eta zein motatakoa den ere zehaztu behar da. Identifikadorea ere jar daiteke, baina ez da beharrezkoa.

[5.1](#) irudiak orokorrean dokumentuek izan behar duten formatua azaltzen du. [5.2](#) irudian gure dokumentu bateko adibide bat ikus daiteke. Lauki berdez datuak agertzen dira, eta,

```

{ "index": { "_index": "indizearen_izena", "_type": "mota", "_id": "identifikadorea" } }
{ "eremua": "balioa" }
{ "index": { "_index": "indizearen_izena", "_type": "mota", "_id": "identifikadorea1" } }
{ "eremua": "balioa" }
{ "index": { "_index": "indizearen_izena", "_type": "mota", "_id": "identifikadorea2" } }
{ "eremua": "balioa" }
...
Ondorengo lerroan datozen datuak indexatu nahi ditugula esan nahi du. Honen
ordez "delete", "create" eta "update" balioak ere onartzen dira, horregatik zehaztu
behar da.

Datuak lerro horretan sartu behar dira.

```

5.1 Irudia: Dokumentuek izan behar duten egitura zehatza.

berez, horrek guztiak lerro bakar batean egon beharko luke. Datuak ondo ikus daitezen jarri dugu modu horretan.

Laburbilduz, 4.1 irudiko formatutik 5.2 irudiko formatura pasatzeko script bat programatu behar izan zen Javaz.

```

{"index":{"_index":"dokumentuak", "type":"hitza"}}
{"analisia":["/hau/hau/DET/M/ABS/ERKARR/S//////////o_z
//hau//edun/ADL//////////HARK//HI/A1//////////o_ez
//hau//ukan/ADT//////////HARK//HI/A1/PNT//////////o_ez",
"aurrekoak1":["/herrialde/herrialde/IZE/MG/ABS/ARR//////////o_z
//herrialde/herrialde/IZE/MG/ABS/ARR//////////o_ez",
"aurrekoak2":["/dituen//ukan/ADT//////////MOS//HARK//HAIEK/A1/PNT//////////o_z
//dituen//edun/ADL//////////ERLT//HARK//HAIEK/A1/PNT//////////o_ez
//dituen//edun/ADL//////////ZHG//HARK//HAIEK/A1/PNT//////////o_ez
//dituen//edun/ADL//////////MOS//HARK//HAIEK/A1/PNT//////////o_ez
//dituen//ukan/ADT//////////ERLT//HARK//HAIEK/A1/PNT//////////o_ez
//dituen//ukan/ADT//////////ZHG//HARK//HAIEK/A1/PNT//////////o_ez",
"aurrekoak3":["/aintzira/aintzira/IZE/M/ABS/ARR/S//////////o_z
//aintzira/aintziratu/ADI//SIN//////////aintzira//DU//AD0IN//////////o_ez
//aintzira/aintzira/IZE/ARR//////////o_ez
//aintzira/aintzira/IZE/MG/ABS/ARR//////////o_ez",
"aurrekoak4":["/188.000/188.000/DET//DZH//////////P//////////o_z
//188.000/188.000/DET/MG/ABS/DZH//////////P//////////o_ez
//188.000/188.000/IZE//ZKI//////////o_ez
//188.000/188.000/IZE/MG/ABS/ZKI//////////o_ez
//188.000/188.000/IZE/M/ABS/ZKI/S//////////o_ez",
"aurrekoak5":["/eta/eta/LOT//////////JNT//////////EMEN//////////o_z
//eta/eta/LOT//////////MEN//////////KAUS//////////AM//////////o_ez",
"hurrengoak1":["/errepublika/errepublika/IZE//ARR//////////o_z
//errepublika/errepublika/IZE/MG/ABS/ARR//////////o_ez
//errepublika/errepublika/IZE/M/ABS/ARR/S//////////o_ez",
"hurrengoak2":["/independientea/independiente/ADJ/M/ABS/ARR/S//////////o_z",
"hurrengoak3":["/da/izan/ADT//////////HUURA/A1/PNT//////////o_z
//da/izan/ADL//////////HUURA/A1/PNT//////////o_ez",
"hurrengoak4":["/1917az/1917/IZE/M/INS/ZKI/S//////////o_z",
"hurrengoak5":["/geroztik/geroztik/ADB//////////ARR//////////o_z",
"aurrekoak":["Konifero baso erraldoiak eta 188.000 aintzira dituen herrialde ",
"ondokoa":["errepublika independentea da 1917az geroztik. Suedia eta ..."]

```

5.2 Irudia: Dokumentu baten egitura.

5.2 Datuen indexazioa

Behin datuak formatu egokira pasatuta, dokumentu horiek automatikoki indexatzeko script bat garatu behar izan da Javaz. Probak egiteko datu asko sartu beharko genituela bagenekienez, ez zen bideragarria banan-banan indexatu behar izatea.

Dokumentu bat indexatzeko terminalean exekutatu beharreko agindua [5.3](#) irudian agertzen dena da

```
curl -XPUT localhost:9200/_bulk --data-binary @dokumentuarenIzena
```

5.3 Irudia: Dokumentuak indexatzeko exekutatu beharreko agindua.

Hori jakinda eta 10.000 dokumentu indexatu beharko genituela kontuan hartuta, ondoak izan ziren jarraitu beharreko bi pausoak:

- Fitxategi batean dokumentu bakoitzeko [5.3](#) irudiko agindua idatzi (bakoitza lerro batean).
- Terminalean fitxategia exekutagarri bat dela zehazteko, eta hori bera exekutatze aginduak exekutatu.

Hori egin ondoren, dokumentu guztiak automatikoki indexatzea lortzen da.

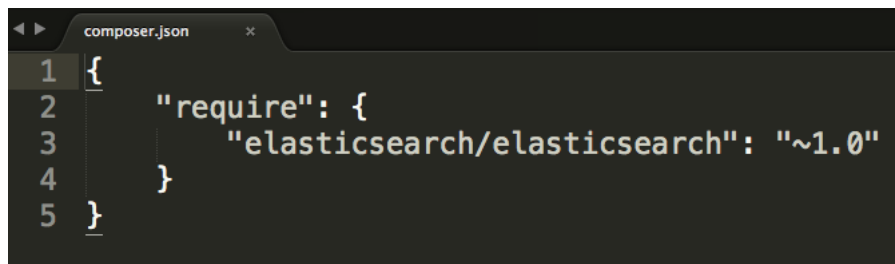
5.3 Elasticsearch-PHP konfigurazioa

Elasticsearch gure web-aplikazioan martxan jartzeko bi baldintza bete behar ziren:

- PHPko 5.4.0 edo bertsio berriago bat.
- [Composer](#). PHPn dependentziak kudeatzeko erabiltzen den tresna da.

Composeren instalazioa

Aplikaziorako behar ditugun liburutegiak deklaratzeko aukera ematen du, eta horren ondoren berak instalatzen ditu. Esan beharra dago *Composer* tresnak PHPko 5.3.2 bertsioa edo berriago bat behar duela.



```
1 {
2     "require": {
3         "elasticsearch/elasticsearch": "~1.0"
4     }
5 }
```

5.4 Irudia: Elasticsearch dependentzia.

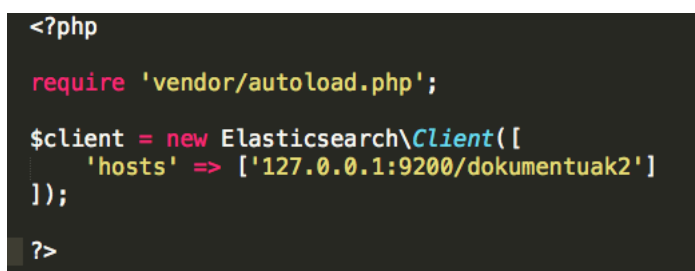
Lehendabizi Elasticsearch dependentzia ezarri behar zen. Horretarako, *composer.json* fitxategi bat sortu zen 5.4 irudian ikus daitekeen edukiarekin.

Ondoren, instalazioa egiteko terminalean beheko bi komandoak exekutatzearekin nahikoa izan zen.

- `curl -s http://getcomposer.org/installer | php`
- `php composer.phar install --no-dev`

Lehenengo komandoak *composer.phar* PHP paketea deskargatzen du, eta bigarrenak instalazioa burutu. Horren ondoren *Composerek* behar izan ditugun dependentzia guztiak instalatu ditu.

Azkenik, bezero berri bat instantziatu behar izan zen (5.5 irudian ikus daitekeen moduan).



```
<?php
require 'vendor/autoload.php';

$client = new Elasticsearch\Client([
    'hosts' => ['127.0.0.1:9200/dokumentuak2']
]);

?>
```

5.5 Irudia: Bezeroa sortzen.

5.4 Web-aplikazioaren ezaugarri eta erabilpen-kasuak

Jarraian, web-aplikazioak dituen hainbat ezaugarri eta erabilpen-kasu azalduko ditugu. Ezaugarrien kasuan, egitura, itxura eta formularioari buruzko azalpenak emango dira.

Erabilpen-kasuei dagokienez, hiru daude: *bilaketa egin/emaitzak ikusi*, *analisiak ikusi* eta *orritz aldatu*.

5.4.1 Aplikazioaren egitura eta itxura

Hasieratik jakin dugu aplikazioaren egiturak nolakoa izan behar zuen, planifikazioko betekizunen atalean zehaztuta dagoen moduan. Jakina da HTML (*HyperText Markup Language*) web-orri baten egitura sortzeko erabiltzen dela, eta guk ere hori erabili beharra izan dugu interfazearen egitura definitzeko: goiburua, gorputza eta oina. 5.6 irudian agertzen da garatu den egitura.

```
<!DOCTYPE html>
<html lang="eu">
  <head>
    <!-- ... -->
  </head>

  <body onload="hasi()">
    <div id="gorputza">
      <header id="goiburua">
        <!-- ... -->
      </header>
      <nav id="menua">
        <!-- ... -->
      </nav>
      <form name="form1" action="index.php" method="post">
        <section id="section">
          <div id="sectionEmaitzak"> <!-- ... --> </div>
          <div id="sectionAnalisia" style="display:none;" <!-- ... --> </div>
        </section>
        <aside id="zutabea">
          <!-- Lehenengo hitza -->
          <fieldset> <!-- ... --> </fieldset>

          <!-- Bigarren hitza -->
          <fieldset> <!-- ... --> </fieldset>

          <!-- Hirugarren hitza -->
          <fieldset> <!-- ... --> </fieldset>

          <input type="submit" value="Bilatu" />
          <input type="submit" value="Garbitu" onclick="atzera(<?php echo 0 ?>)" />
        </aside>
      </form>
      <footer id="oina"></footer>
    </div>
  </body>
</html>
```

5.6 Irudia: Web-aplikazioaren interfazearen egitura (HTML kodea).

Itxurari dagokionez, Ixa Taldearen koloreak errespetatzen saiatu gara eta, horrez gain, erabiltzaileari begira interfaze txukuna sortzen. Horretarako CSS teknologia erabili da. Atal honetan esan beharra dago itxura guztia ez duela ikasleak garatu. Itxura hobea lortzeko Interneten eskuragarri dauden baliabideak erabili ditu, jarraipen eta kontrolean adierazi den moduan. Gauzak horrela, denbora gutxian itxura txukun bat egitea lortu da.

5.4.2 Formularioa

Diseinuaren atalean esan genuen formularioko datuak JSON formatuko fitxategian gorde-ta zeudela, eta formularioa sortzeko datuak bertatik atzitu beharko zirela. 5.7 irudian ikus daiteke erabiltzaileak hautatutako kategoriaren arabera nola sortzen diren formularioko ezaugarriak, sasi-kodean idatzita ulergarriagoa izan dadin.

Lehendabizi, aurretik beste kategoriren baten ezaugarriak sortuta baldin badaude ezabatu egiten dira. Ondoren, *getJSON()* funtzioaren bidez fitxategia atzitzen da. Hautatutako kategoriaren ezaugarrien izenak hartzen dira ondoren, eta horiekin HTMLko *select* etiketak sortzen dira. Azkenik, ezaugarri bakoitzak barruan daturen bat duen edo ez begirutzen da. Ezezko kasuan, *select* motako etiketa izan beharrean, *input* motako etiketa izan beharko da erabiltzaileak nahi duena idatz dezan. Baiezko kasuan, berriz, *select* motako etiketari balio posible horiek gehitzen zaizkio.

Bestalde, *select*ari balioak gehitzerakoan, aurretik zein balio aukeratuta zegoen begirutzen da lehendabizi. Interfazean orrialdez aldatzerakoan formularioa zegoen bezala mantendu behar da datuak berdin-berdin mantentzeko eta erabiltzailearen bilaketa ez aldatzeko. Horregatik, zati hori nolabait kontrolatu beharra zegoen.

Beste kategorია guztiekin ere gauza bera egiten da.

```
//Aurretik beste ezaugarriren bat baldin bazegoen, ezabatu.
//Ezaugarrien izenak gordetzeko array bat hasieratu:
var ezaug = new Array();

getJSON('datuak.json' fitxategia atzitu){
  for (Hautatutako kategoriaren ezaugarrien izen bakoitzeko (val)){
    //Izen bakoitza ('val' parametroa) array batean gorde.
    ezaug[i] = val;
    //html etiketa sortu:
    '<select id="'+val+'" name="'+val+'"></select>'
    //lehenengo aukera sortu, ezaugarriaren izenarekin:
    '<option value="'+ize[i+1]+'"'>'+val+'</option>'
  }
  for (Hautatutako kategoriaren ezaugarri bakoitzeko (val)){
    if (ezaugarriak barruan elementurik ez badu){
      //ezabatu aurrez sortu den etiketa.
      //beste html etiketa mota bat sortu, input motakoa, erabiltzaileak balioa eskuz sartu ahal izateko:
      '<input type="text" id="'+ezaug[i]+'"' name="'+ezaug[i]+'"' placeholder="—'+ezaug[i]+'—"></input>'
    }else{
      for (ezaugarriaren elementu bakoitzeko){
        if (aurrez hautatua baldin badago){
          '<option selected="true" value="'+val[ezaug[i]][t]['balioa']+'"'>'+val[ezaug[i]][t]['izena']+'</option>'
        }else{
          '<option value="'+val[ezaug[i]][t]['balioa']+'"'>'+val[ezaug[i]][t]['izena']+'</option>'
        }
      }
    }
  }
}
```

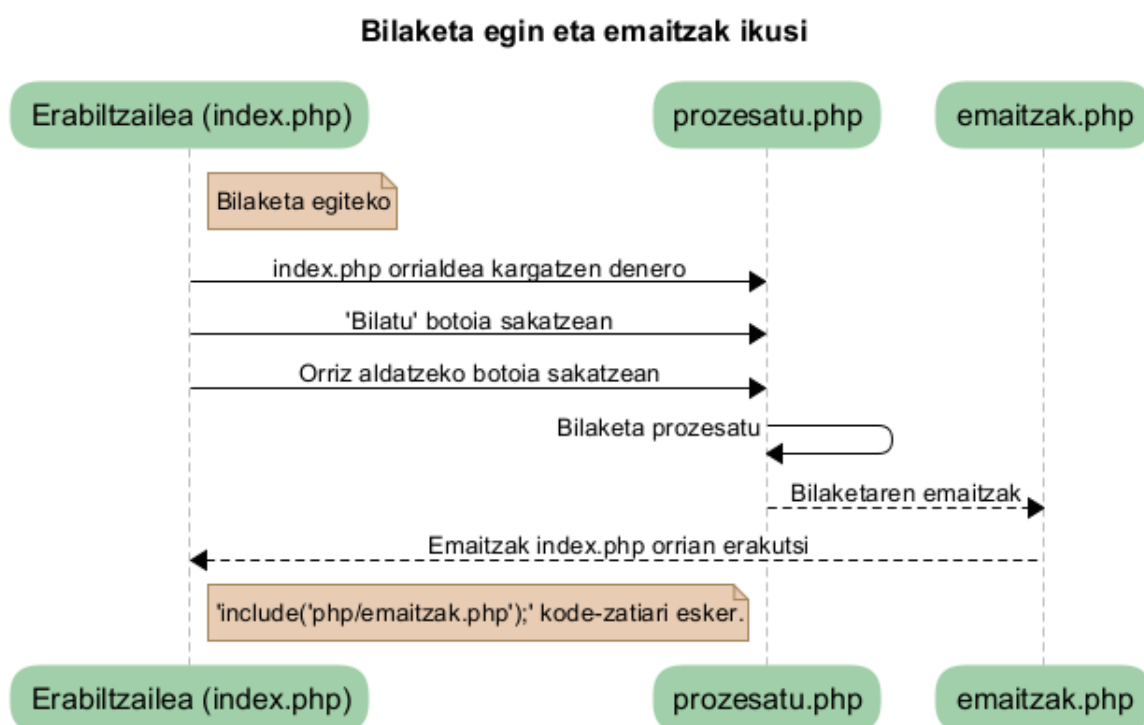
5.7 Irudia: Formularioaren sorkuntza (sasi-kodea).

5.4.3 Bilaketa egin/emaitzak ikusi erabilpen-kasua

Sekuentzia-diagrama

Hasteko, garrantzitsua da esatea *prozesatu.php* eta *emaitzak.php* orrialdeak *include* aginduaren bidez txertatuta daudela *index.php* orrian. Horrek esan nahi du, *index* orria kargatzen den guztietan (bilaketa botoiari ematen zaionean edota orriz aldatzeko botoia sakatzen denean) beste biak ere kargatzen direla.

Hori azalduta, sekuentzia-diagrama honek ez du misterio handirik. *Index* orrian bilaketa bat egiten denean *prozesatu* orrian prozesatzen da, eta *emaitzak* orrian emaitzak idazten dira.



5.8 Irudia: Bilaketa egin/emaitzak ikusi erabilpen-kasuaren sekuentzia-diagrama.

Garapena

Bilaketak egiteko, lehendabizi formulariotik jasotako PHP parametroak gorde behar dira eta, ondoren, parametro horiekin hiru hitzei dagozkien *wildcard* espresioak osatu. Behin horiek osatuta, erabiltzaileak bilaketa zenbat hitzekin izatea nahi duen kontrolatzen da (espresioaren bat hutsik baldin badago, hitz hori ez duela bilatu nahi estimatzen da).


```
</br>
<?php
$results = $query['hits']['hits'];
if(isset($results)){
    // $j = 0;
    foreach ($results as $r) {
        // $j++;
        $anali = $r['_source']['analisa'];
        $a = split(" ", $anali);
        $analis = split("\\\\", $anali);
        $analisiarenForma = $analis[1];

        // Hitz konposatua bada, '%' ikurra kendu.
        if (strpos($analisiarenForma, '%') === true && (substr($analisiarenForma, 0, 1) != '%')){
            $analisiarenForma = str_replace('%', ' ', $analisiarenForma);
        }

        <div class="result">
            <?php
                echo $r['_source']['aurrekoa'];
            ?>
            <a onclick='analisaAktibatu("<?php echo $anali; ?>"')> <?php echo $analisiarenForma; ?></a>
            <?php
                echo $r['_source']['ondokoa'];
            ?>
        </div>
    }
}
?>
</br>
```

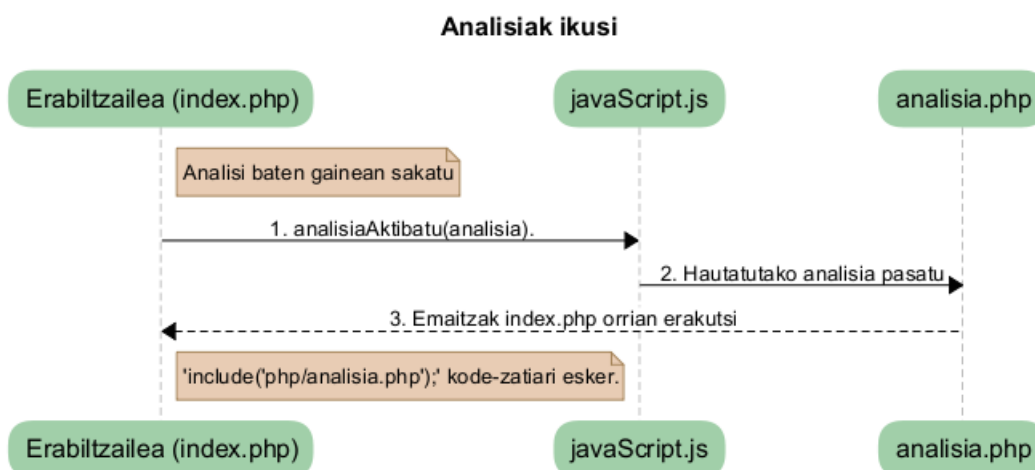
5.10 Irudia: Emaitzak nola pantailaratu erakusten duen kode zatia.

5.4.4 *Analisiak ikusi* erabilpen-kasua

Sekuentzia-diagrama

Aurreko erabilpen-kasuan azaldu den moduan, hemen ere azpimarratu behar da *analisiak.php* orrialdea *include* aginduaren bidez *index.php* orrian txertaturik dagoela.

Bilaketa bat egin denean eta emaitzak pantailan erakutsi direnean, bilatutako hitza sartzeko aukera du erabiltzaileak. Hitz horretan klikatuz, *analisiakAktibatu()* *JavaScript* funtzioari deitzen zaio, hitzaren *analisiak* parametro moduan pasatuz. Funtzio horrek, POST moduko eskaera bat egiten dio zerbitzariari (*analisiak.php* orrialdeari), *Ajax* bidez, *analisiak* pantailaratzeko.



5.11 Irudia: *Analisiak ikusi* erabilpen-kasuaren sekuentzia diagrama.

Garapena

Aplikazioan bilaketa bat egiten denean emaitzak agertzen dira, baina emaitzetako hitz batean klikatzen bada, hitz horren *analisiak* ikusten dira. Azkeneko pauso hori egiteko *Ajax* teknologia erabili da. 5.12 irudian ikus daiteke kodea.

Analisiak erakusteko 5.10 irudian ikus daiteke nola deitzen zaion *analisiakAktibatu* funtzioari. Bertan, *Ajax* bidez *analisiak.php* fitxategi bati POST bidez aukeratutako hitzaren *analisiak* pasatzen zaiola ageri da, eta, azkenik, emaitza idazten dela. *analisiak.php* fitxategiak *analisiak* dagoen formatutik ezaugarri guztiak lortzen ditu, eta, ondoren, hutsik ez daudenak pantailaratzten ditu. Azkeneko orrialde horretan, 5.13 irudiak erakusten du *analisiak* nola jasotzen den.

```
function analisisAktibatu(a){  
    $.post("php/analisis.php",{ "analisis":a },function(erantzuna){  
        $("#div_session_write").html(erantzuna);  
    });  
  
    document.getElementById("sectionEmaitzak").style.display='none';  
    document.getElementById("sectionAnalisis").style.display='inline';  
    document.getElementById("an").style.fontWeight = "bold";  
    document.getElementById("em").style.fontWeight = "normal";  
}
```

5.12 Irudia: Analisisak (Ajax kodea).

```
<?php  
  
$analisiak = $_POST['analisis'];
```

5.13 Irudia: *analisis.php* orrian analisisa nola jaso.

5.4.5 Orriz aldatu erabilpen-kasua

Erabilpen-kasu honen sekuentzia-diagrama bilaketa egiteko erabilpen-kasuaren berdina denez, ez dugu berririo azalduko.

Orrialdeak kontrolatzeko `hidden` motako HTML etiketa bat sortu zen eta bere balioa aldatuz lortu genuen orrialdeen aldaketa egitea. 5.14 kodean agertzen den moduan, *aurrera* botoiari ematen zaionean `page` identifikadorea duen etiketaren balioari bat gehitzen zaio. Atzera joateko botoiari ematen zaionean, berriz, balioari bat kentzen zaio.

```
function atzera(p){  
    var page = p;  
    var page = page-1;  
  
    document.getElementById('page').value = page;  
}  
  
function aurrera(p){  
    var page = p;  
    var page = page+1;  
  
    document.getElementById('page').value = page;  
}
```

5.14 Irudia: Orrialde aldaketa (JavaScript kodea).

5.5 Erabilitako teknologiak

Web-aplikazioa garatzeko hainbat teknologia erabili dira, eta horietako batzuk dagoeneko agertu dira aurretik, baina atal honetan gehiago zehaztuko ditugu:

- **HTML**: Web-aplikazio baten egitura sortzeko erabili da.
- **CSS3**: Aplikazioari itxura hobe emateko erabili da.
- **JavaScript**: JavaScript web-orri dinamikoak sortzeko zuzenduta dagoen lengoaia da, bezeroaren aldekoa. Guk hainbat funtzionalitate egiteko erabili dugu:
 - Formularioko ezaugarriak kategoriaren arabera ezberdintzeko.
 - Emaizena atala aktibatu edo desaktibatzeke.
 - Analisisien atala aktibatu edo desaktibatzeke.
 - Orrialdeak kontrolatzeko.
- **jQuery**: Formularioko datuak kanpoko JSON fitxategi batetik atzitzeko erabili da.
- **Ajax**: Analisisiak ikusteko erabili da.
- **Elasticsearch-PHP**: Bilaketak egiteko erabili da motor hau.

6. KAPITULUA

Esperimentuak

Kapitulu honetan egindako esperimentu batzuk azalduko dira aplikazioaren erantzun-denborak aztertzeko.

Lau bilaketa mota egin dira:

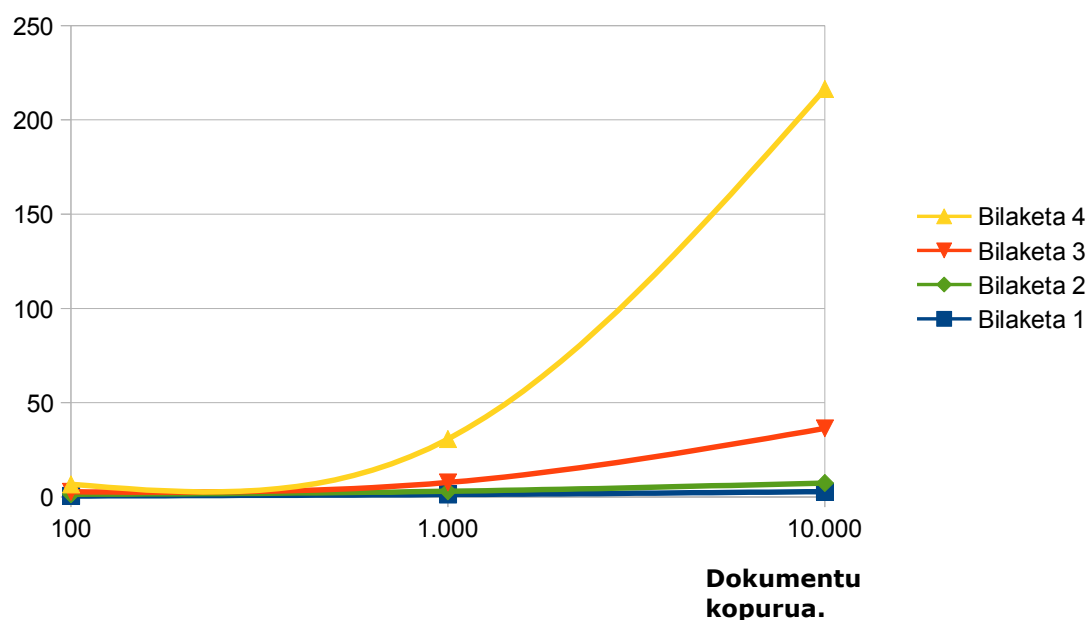
- **Bakarra:** Bilaketa hitz bakarrarekin. Hemen bi bereizi dira: emaitza gutxiko bilaketa eta emaitza askoko beste bat.
 - **Emitza gutxikoa:** Lema "denbora"duten hitz guztiak bilatu.
 - **Emitza askokoa:** Izena diren hitz guztiak bilatu (Kategoria="izena").
- **Konposatua** Bilaketa bi hitzekin. Hemen ere bi zatitan banatu dira: distantzia 1 izanik eta distantzia 5 izanik.
 - **Distantzia 1:** Lehenengo hitzaren lema "urrats"izan beharko da eta bigarrenaren kategoria "aditza"izan beharko da. Bien arteko distantzia 1 izatea nahi da, eta aurretik edo atzetik. Bilaketa honen bidez "urrats"hitzen ondoan zein aditz erabiltzen den gehien ikus daiteke.
 - **Distantzia 5:** Bilaketa arruntaren berdina, baina distantzia 5 izanik.

6.1 irudian agertzen den grafikoan ikus daitezke 6.1 taulan dauden erantzun-denborak, indexatutako dokumentu kopuruaren arabera. X ardatzean ikus daitezkeen moduan, 100, 1.000 eta 10.000 dokumenturekin egin dira probak. Taulan bilaketa bakoitzak itzuli dituen emaitza kopuruak ere sartu ditugu ondorio sakonagoak atera ahal izateko.

Dokumentu kopurua	Bilaketa 1		Bilaketa 2		Bilaketa 3		Bilaketa 4	
	Denb	E kop	Denb	E kop	Denb	E kop	Denb	E kop
100	0.51s	70	0.9s	58400	1.3s	5	4s	7
1.000	1.3s	538	1.7s	461765	4.7s	38	23s	64
10.000	2.8s	1804	4.6s	1762983	29s	151	180s	150

6.1 Taula: Aplikazioaren erantzun-denborak eta emaitza kopuruak.

**Erantzun-denbora
segundotan.**



6.1 Irudia: Aplikazioaren erantzun-denborak dokumentu kopuruaren arabera.

Lehendabizi bilaketa bakoitza zeini dagokion azalduko dugu:

- **Bilaketa 1:** Bakarra + emaitza gutxi.
- **Bilaketa 2:** Bakarra + emaitza asko.
- **Bilaketa 3:** Konposatua + distantzia 1.
- **Bilaketa 4:** Konposatua + distantzia 5.

100 dokumenturekin ari garenean, ikus daiteke nahiz eta bilaketaren konplexutasuna igo, ez duela eragin handirik erantzun-denboran, oso onargarria baita irauten duena. 1.000

dokumenturekin igoera txiki bat nabari da laugarren bilaketan (konplexuena), baina beste bilaketen denborak onargarriak direla esan daiteke. 10.000 dokumenturekin oso argi ikusten da dokumentu kopuruak eta bilaketen konplexutasunak duten garrantzia edo eragina. Lehenengo hiru bilaketen erantzun-denborak nahiko egokiak dira, baina laugarrenarekin asko irauten du.

Horretaz aparte, bilaketa bakoitzaren emaitzak ere jaso ditugu ea erantzun-denboran eragina duen aztertzeko. Baina taula ikusita argi dago horrek ez duela eraginik. Horren adibide argi bat jartzearen, 10.000 dokumenturekin bigarren bilaketa egiteko 4,6 segundo behar izan dira, emaitza 1.762.983 izanik. Dokumentu kopuru berdinarekin laugarren bilaketa egiteko, berriz, 180 segundo behar izan dira emaitza 250 izanik.

Aurreko guztia laburbilduz, esan daiteke hitzen arteko distantzia handitzeak eragin handia duela aplikazioaren denboran.

7. KAPITULUA

Jarraipena eta Kontrola

Atal honetan proiektuan egindakoa planifikatutakoarekin alderatuko dugu eta gertatu diren gorabeherei zergatia eman. Jarraipena eta kontrola egiteak uste baino lan gehiago eskatu du, izan ere, planifikazioa zati handi batean aldatu egin behar izan da. Aldaketa horiek guztiak banan-banan azalduko dira kapitulu honetan.

Lehendabizi gertaturiko desbideraketekin hasiko gara, proiektuko aldaketa nagusien zergatia uler dezagun. Jarraian, aldaketa horiek mugarrietan, irismen-mailan, kalitate-planean eta arriskueta izan duten eragina azalduko da.

7.1 Desbideraketak

Onerako edo txarrerako, ataza guztietan gertatu dira desbideraketak. [7.1](#) taulan agertzen dira estimatutako eta benetako denborak.

Kudeaketa

- **1.1:** Planifikazioa egiteko uste baino denbora gutxiago behar izan da, baina ez da ezer bereziagatik izan.
- **1.2:** Jarraipena burutzeko, berriz, denbora gehiago behar izan da. Hasteko, uste baino bilera gehiago egin behar izan direlako. Bestalde, plangintzak jasan dituen aldaketek ere eragin dute denbora handitzea. Hala ere, 3 ordu gehiago besterik ez dira behar izan.

ID	Ataza	Estimazioa	Burututakoa	Desbideraketa
Kudeaketa				
1.1	Plangintza	25 ordu	22 ordu	-3 ordu
1.2	Jarraipena	13 ordu	16 ordu	+3 ordu
1.3	Kontrola	14 ordu	12 ordu	-2 ordu
		52 ordu	50 ordu	-2 ordu
Ikerkuntza				
2.1	Elasticsearch	60 ordu	54 ordu	-6 ordu
		60 ordu	54 ordu	-6 ordu
Diseinua				
3.1	Datu-egitura	90 ordu	82 ordu	-8 ordu
3.2	Web-aplikazioaren interfazea	2 ordu	1 ordu	-1 ordu
		92 ordu	83 ordu	-9 ordu
Garapena				
4.1	Lan-ingurunea	6 ordu	8 ordu	+2 ordu
4.2	Beharrezko script-ak	15 ordu	19 ordu	+4 ordu
4.3	Web-aplikazioa	65 ordu	112 ordu	+47 ordu
4.4	Probak	2 ordu	1 ordu	-1 ordu
		88 ordu	140 ordu	+52 ordu
Dokumentazioa				
5.1	Memoria	92 ordu	108 ordu	+16 ordu
5.2	Aurkezpena	8 ordu	7 ordu	-1 ordu
		100 ordu	115 ordu	+15 ordu
GUZTIRA: 442 ordu				

7.1 Taula: Desbideraketak.

- **1.3:** Kontrola egiteko estimatutakoa baino bi ordu gutxiago behar izan dira, baina, planifikazioan gertatu den moduan, ez da ezer bereziagatik izan.

Ikerkuntza

- **2.1:** Ordu kopuruei dagokienez, Elasticsearch ikasten gertatu den desbiderapena onerako izan da. Kode irekikoa denez, [web-orri](#) ofizialean dago behar den informazio guztia adibideekin batera. Bestalde, zuzendariak EHUKo liburutegian eskatutako liburua ere eduki du ikasleak eskuragarri. Horregatik behar izan da denbora gutxiago hau egiteko. Hala ere, aurrerago ikusiko den moduan, denboran zehar atzerapena gertatu da Urtarrileko azterketak zirela medio.

Diseinua

- **3.1:** Datu-egituraren diseinua egiteko ere denbora gutxiago behar izan da, baina ez da desbideraketa handia izan. 4. orrialdean zehaztutako betekizunak beteko zituen diseinu bat behar zen, eta ez da lan erraza izan, ahalik eta diseinu optimizatuena lortu nahi baitzen web-aplikazioaren eraginkortasuna hobetzeko.
- **3.2:** Web-aplikazioaren interfazea diseinatzeko ordu bat gutxiago besterik ez da behar izan, interesatuek oso argi baitzekiten nola egin.

Garapena

- **4.1:** Lan-ingurunea prestatzeko uste baino pixka bat gehiago behar izan da, gertaturiko erroreak direla eta.
- **4.2:** Beharrezko script-ak sortzeko ere pixka bat gehiago behar izan da. Desbiderapen hau, batik bat, hasieran datuek zein egitura jarraitu beharko zuten ez genekielako gertatu da. Zuzendariek ikasleari pasatutako datuak formatuz aldatzeko ezinbestekoa zen egitura berria nolakoa zen jakitea.
- **4.3:** Web-aplikazioa garatzean oso desbiderapen handia gertatu da. Hasieran, aplikazioa garatzeko Elasticsearch-ek gomendatzen duen Kibana interfazea erabiltzea pentsatu genuen, denbora murriztuko genuelakoan. Baina, hala ere, bertan formularioa txertatu behar zen eta ikaslea hori egiten hasi zenean ez zela bideragarria ikusi zuen. Hortaz, zuzendariekin bilera bat egin ondoren, tresnaz aldatzeko erabakia hartu zen, azken finean helburua datuen berreskurapena erakutsiko zuen prototipo bat garatzea baitzen. Aldaketa horrek atzerapen handia ekarriko zuenez, ikasleak dagoeneko ezagutzen zuen tresna erabiltzea erabaki zen: PHP. Hori izan da desbiderapen handi honen arrazoia.
- **4.4:** Bilaketa mota guztiak ondo dabiltzala ziurtatzeko espero baino ordu bat gutxiago behar izan da.

Dokumentazioa

- **5.1:** Planifikazioak jasan dituen aldaketen eraginez, gehienbat, memoriak ere desbideraketa garrantzitsua jasan du, aldaketak eta horien arrazoiak ondo azaldu behar baitziren. Guztira 16 ordu gehiago behar izan dira.

- **5.2:** Defentsa-egunerako beharko den aurkezpena sortzeko uste baino gutxiago behar izan da, baina ez da desbiderapen handia izan.

7.2 Gantt diagrama

Aurrez azaldutako desbiderapenen ondorioz egutegia aldatu behar izan da. Beraz, hori modu grafikoan erakutsiko duen Gantt diagrama berria sortu da (ikus [7.1](#)).

Azpimarratzekoa da ikerkuntza-atza burutzerakoan gertatu den atzerapena. Planifikazioa egin zenean, nahiz eta azterketak izan, honi denbora bat eskainiko zitzaiola pentsatu zen (nahiz eta denbora txikia izan). Baina hori ez zen horrela izan, eta azterketetan zehar proiektua alde batera utzi zen. Atzerapen horrek diseinuari ere eragin dio.

Horretaz aparte, web-aplikazioaren garapenean ere atzerapen nabarmena gertatu da, irudian argi ikus daitekeen moduan, desbiderapenetan azaldu diren arrazoiengatik.

7.3 Mugarriak

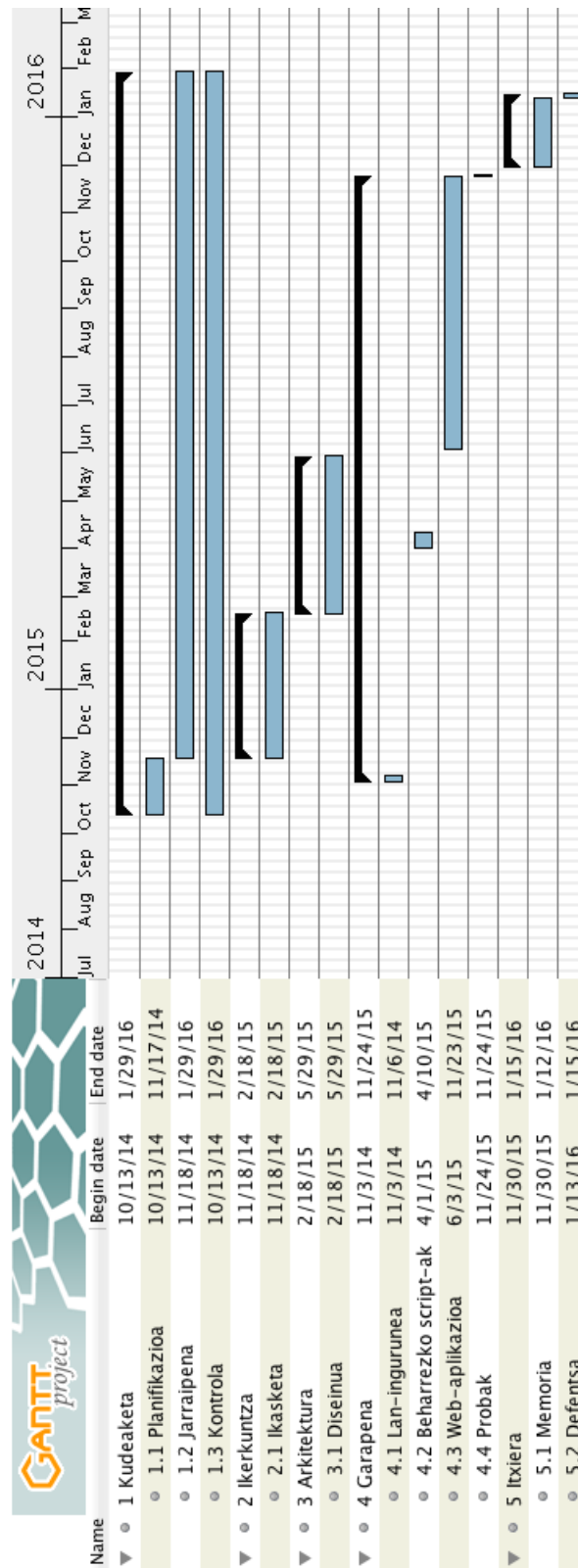
Gertatutako atzerapenen ondorioz, eta interfazea Kibana erabiliz egitea bideragarria ez zela konturatu ginenean, mugarriak aldatu behar izan genituen. Proiektua defendatzeko epeak azaroan eta otsailean ziren, eta PHPrekin aurrera jarraitzea erabaki zenean iraila zenez, ez zen bideragarria azarorako dena prest egongo zela pentsatzea. Hortaz, proiektuaren defentsa otsailera arte atzeratzea erabaki zen.

Hori dela eta, honakoak izan ziren aldatu genituen mugarriak:

- **2015-12-15:** Web-aplikazioaren garapena bukatu.
- **2016-01-20:** Memoria bukatu.
- **2016-01-30:** Defentsa-eguneko aurkezpena bukatu.

7.4 Irismen-maila

[5.](#) orrialdean bi irismen-maila zehaztu genituen, eta denboraren arabera maila batera edo bestera iritsiko ginela zehaztu. Gauzak horrela, eta mugarriak atzeratu behar izan ditugula kontuan hartuz, lehenengo mailara iritsi gara, oinarritzokora.



7.1 Irudia: Gantt diagrama.

7.5 Kalitate-plana

12. orrialdean burutu zen kalitate-planean dauden irizpideak banan-banan aztertuko ditugu, egindako proiektuarekin alderatuz:

- **Denboraren kalitatea:** Ez dira planifikatutako denborak bete. Orokorrean nahiko desbiderapen handia gertatu da.
- **Emaitzaren kalitatea:** Emaitzaren kalitatea guztiz bete da. Web-aplikazioaren interfazea erabiltzailearen aldetik erabilerraza da eta, gainera, erantzun-denbora murrizteko saiakera arrakastatsua izan da.
- **Kodearen kalitatea:** Kodea ahalik eta txukunen eta optimizatuen utzi da.
- **Dokumentazioaren kalitatea:** Garrantzia handia eman zaio dokumentazioari, izan ere, gauza asko zeuden azaltzeko.

7.6 Eskuraketak

Planifikazioa egin genuenean ez genuen adierazi interneten eskuragarri dauden kode zatiak erabiliko genituela eta azkenean horrelako eskuraketak ere egin dira.

- Formularioaren itxura hobetzeko. Bisitatu [orrialde](#) hau.
- Animazioa duten botoiak jartzeko. Bisitatu [orrialde](#) hau.

7.7 Arriskuak

13. orrialdean azaldutako arriskuei dagokienez, A1 eta A2 arriskuak ez gertatzeko kudeaketa egokia egin da. Esan bezala, kontu handiz ibili gara horiek gerta ez zitezen. Zorritzarrez, A3 arriskua ezin izan da ekidin, eta horrek epeak atzeratzeko ondorioa ekarri du.

Bada kudeatu ez zen arrisku bat eta gertatu dena. Urtarrileko azterketek irauten zuten denboran ez zen aurreikusi ikasleak proiektua alde batera uzteko zuen arriskua, eta horrek eragina izan du atzerapenean.

8. KAPITULUA

Ondorioak

Helburuak orokorrean bete dira, nahiz eta denboran desbiderapenak gertatu diren. Irismen-mailak zehaztu genituenean argi utzi genuen oinarritzko mailara iristea beharrezkoa izango zela, eta hori da egin dena. Web-aplikazio bat garatu da, linguistikoki anotatutako corpus batean bilaketak egiteko; horretarako, *IR*ren arloan erabiltzen diren teknikak erabili dira, eta ongi eskalatzen duten aplikazioak sortzeko egokia den tresna bat *Elasticsearch* erabiliz.

Proiektua Software Ingeniaritza alorrekoa izanik, programazio aldeko ezagutzak gehiago sakondu dira, baina ez hori bakarrik. Proiektu bat aurrera eramateko arazoan aurrean nola jokatzen ere ikasi da. Bagenekien proiektua egiten hasi aurretik ez zela erraza izango eta arazo ugarirekin topatuko ginela bidean zehar, baina, nahiz eta atzerapenak gertatu diren, une larriei aurre egiteko gai izan gara, eta proiektua modu txukun batean bukatu dugu.

Lehendabizi *Elasticsearch*en ikasketarekin hasi gara. Konputagailuan proba txikiak eginenez lortu da horri buruzko ezagutza.

Ondoren, datu-egitura bat diseinatu behar zen. Ikerketan zehar egindako probek atal honetan asko lagundu dute. Prozesu honetan zuzendarien eta ikaslearen arteko bilerek izugarritzko garrantzia izan dute, zuzendariak esan baitiote ikasleari hurrengo bilerarako zer lortu. Horrela, pausoz pauso, erabiltzaileari eman nahi genizkion aukera guztiak eskaintzeko gai den datu-egitura bat osatzea lortu da. Aipatzekoa da, hemen, hitzen analisisiek duten formatuak (Ixa Taldeak definitutakoa, zuzendariak ikasleari proposatutakoa) erabat baldintzatu dutela erabili dugun datu-egitura.

Hurrengo pausoa web-aplikazioa garatzea izan da. Esan bezala, corpus oso handien gai-

nean bilaketa aurreratuak egiteko aplikazio bat sortu da. Erantzun-denbora murriztearren, eskalagarria den tresna bat erabili da hau egiteko (*Elasticsearch*). Hala ere, ez dugu gure proiektuan horren inguruko esperimenterik egin. Bilaketa arruntak eta konplexuagoak egiteko aukera eman zaio erabiltzaileari, eta emaitzak eta hitzen analisiak modu txukun batean erakusten zaizkio.

8.1 Ikasitako lezioak

Proiektuaren bizi-zikloan zehar orain arte garrantzia handirik ematen ez zitzaion atal batzuen garrantzia ikusi da, eta ikasleari etorkizunean izugarri lagunduko dio denbora honetan ikasi duen guztiak.

- **Eguneko helburu txikiak markatu:** Egon dira aurrerapenik lortu ez diren egunak, eta horrek motibazioa izugarri jaisten du askotan. Proiektu bat dagoeneko martxan dagoenean motibazioa galtzea ikasleari gerta dakioken gauzarik okerrenetariko bat da, eta hori ekiditeko helburu txikiak zehazteak izugarri laguntzen du.
- **Kodea berrerabili:** Interneten eskuragarri dauden eta berrerabil daitezkeen kode zati ugari daude. Horiek bilatzen denbora pasatuz gero, etorkizunean denbora asko aurrezten da. Zertarako ibili gorpila behin eta berriro asmatzen?
- **Bilera-aktak:** Proiektuan zehar hartutako erabaki guztiak zuzendarien eta ikaslearen artean adostutakoak izan dira. Izugarri laguntzen du bileretan hitz egindakoa dokumentatzeak eta ondo gordetzeak, gerora begira informazio hori galdu edo ahaztu egin daiteke eta.
- **Planifikazioa:** Proiektu honetan benetan konturatu gara planifikazioak duen garrantziaz. Hasieratik planifikazioaren atalak dituen puntu guztiak zehazten ez bada, oso zaila izango da proiektuak aurrera egiten duen heinean amaiera finkatzea edota irismen-maila onargarria zein den jakitea. Hortaz, merezi du hasieran behar diren orduak hau egiten pasatzeak.

8.2 Etorkizuneko lanak

Nahiz eta proiektuaren helburuak bete diren, web-aplikazioak baditu hobe daitezkeen bi atal nagusi:

- **Datu-egitura:** Datuen egitura zein izango zen erabaki genuenean bagenekien ez zela modurik onena, informazio asko errepikatzen baitzen. Beraz, pentsatzen dugu interesgarria litzatekeela etorkizunean hobeto ikertzea gai hori, eta ikustea ea lor daitezkeen antzeko emaitzak (eraginkortasun aldetik) datuen errepresentazioan errenduntzia gutxituz.
- **Aplikazioaren ezaugarriak:** Armiarma aplikazioarekin konparatuz gureak ezaugarri gutxiago eskaintzen ditu, analisi berriak sortzeko aukera ematea, adibide bat jartzearen. Beraz, etorkizunean falta diren ezaugarri horiek sar daitezke.

Horietaz aparte, badago etorkizunean egin daitekeen beste lan bat:

- **Eskalagarritasunaren azterketa:** Esan bezala, eskalagarria den tresna da *Elastic-search*, baina gure proiektuan ez dugu horretan murgiltzeko aukera izan. Beraz, interesgarria izango litzateke, hainbat makinatan exekutatzen den kasuan sistemaren eraginkortasuna neurtu ahal izatea. Horrela, bide batez, *sharden* erabilgarritasuna eta eragina argiago ikusiko litzateke.

Hala ere, horiek hobekuntza posible moduan jarri ditugu. Izan ere, Ixa taldean oso interesatuta daude proiektuan landu den bezalako bideak jorratzen, linguistikoki anotatutako corpusak gero eta gehiago erabiltzen baitira, eta gero eta handiagoak baitira.

Eranskinak

1. ERANSKINA

Ezaugarri linguistikoak gordetzen dituen fitxategia.

```

{
  "Kategoriak" : [
    { "balioa" : "IZE",
      "izena" : "Izena",
      "ezaugIzenak" : [ "Azpikategoriak", "Kasua", "Mugatasuna", "Numeroa" ],
      "Ezaugarriak" : [
        { "Azpikategoriak" : [
          { "balioa" : "ARR", "izena" : "Izen arrunta" },
          { "balioa" : "LIB", "izena" : "Leku-izen berezia" },
          { "balioa" : "IZB", "izena" : "Pertsona-izen berezia" },
          { "balioa" : "ZKI", "izena" : "Zenbakia" }
        ] },
        { "Kasua" : [
          { "balioa" : "ABL", "izena" : "Ablatiboa" },
          { "balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa" },
          { "balioa" : "ABS", "izena" : "Absolutiboa" },
          { "balioa" : "ABZ", "izena" : "Adlatibo bide zuzeneko" },
          { "balioa" : "ALA", "izena" : "Adlatiboa" },
          { "balioa" : "BNK", "izena" : "Bankaria" },
          { "balioa" : "DAT", "izena" : "Datiboa" },
          { "balioa" : "DES", "izena" : "Destinatiboa" },
          { "balioa" : "DESK", "izena" : "Deskribatzailea" },
          { "balioa" : "ERG", "izena" : "Ergatiboa" },
          { "balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa" },
          { "balioa" : "GEN", "izena" : "Genitibo edutezkoa" },
          { "balioa" : "INE", "izena" : "Inesiboa" },
          { "balioa" : "INS", "izena" : "Instrumentala" },
          { "balioa" : "MOT", "izena" : "Motibatiboa" },
          { "balioa" : "PAR", "izena" : "Partitiboa" },
          { "balioa" : "PRO", "izena" : "Prolatiboa" },
          { "balioa" : "SOZ", "izena" : "Soziatiboa" }
        ] },
        { "Mugatasuna" : [
          { "balioa" : "M", "izena" : "Mugatua" },
          { "balioa" : "MG", "izena" : "Mugagabea" }
        ] },
        { "Numeroa" : [
          { "balioa" : "S", "izena" : "Singularra" },
          { "balioa" : "P", "izena" : "Plurala" },
          { "balioa" : "PH", "izena" : "Plural hurbila" }
        ] }
      ]
    }
  ]
}

```

```

    ]},
    { "balioa" : "ADI", "izena" : "Aditza",
      "ezaugIzenak" : [ "Azpikategoriak", "AditzMota", "Aditzoina", "Aspektua", "Erlazioa", "Kasua",
        "LaguntzaileMota", "ModuaDenbora", "Numeroa" ],
      "Ezaugarriak" : [
        {"Azpikategoriak" : [
          {"balioa" : "SIN", "izena" : "Aditz sinplea"},
          {"balioa" : "FAK", "izena" : "Aditz faktitiboa"},
          {"balioa" : "ADK", "izena" : "Aditz konposatua"},
          {"balioa" : "ADP", "izena" : "Aditz perifrastikoa"}
        ]},
        {"AditzMota" : [
          {"balioa" : "ADIZE", "izena" : "Aditz-izena"},
          {"balioa" : "ADOIN", "izena" : "Aditzoina"},
          {"balioa" : "PART", "izena" : "Aditz-partizipioa"}
        ]},
        {"Aditzoina" : [
        ]},
        {"Aspektua" : [
          {"balioa" : "GERO", "izena" : "Etorkizuna"},
          {"balioa" : "BURU", "izena" : "Burutua"},
          {"balioa" : "PNT", "izena" : "Puntukaria"},
          {"balioa" : "EZBU", "izena" : "Burutugabea"}
        ]},
        {"Erlazioa" : [
          {"balioa" : "AURK", "izena" : "Aurkaritzakoa"},
          {"balioa" : "BALD", "izena" : "Baldintzazkoa"},
          {"balioa" : "DENB", "izena" : "Denborazkoa"},
          {"balioa" : "EMEN", "izena" : "Emendiozkoa"},
          {"balioa" : "ERLT", "izena" : "Erlatibozkoa"},
          {"balioa" : "ESPL", "izena" : "Esplikatiboa"},
          {"balioa" : "HAUT", "izena" : "Hautakaria"},
          {"balioa" : "HELB", "izena" : "Helburuzkoa"},
          {"balioa" : "KAUS", "izena" : "Kausazkoa"},
          {"balioa" : "KONPL", "izena" : "Konpletiboa"},
          {"balioa" : "KONT", "izena" : "Kontzesiboa"},
          {"balioa" : "MOD", "izena" : "Moduzkoa"},
          {"balioa" : "MOD/DENB", "izena" : "Moduzkoa/Denborazkoa"},
          {"balioa" : "MOS", "izena" : "Menderagailu-osagaia"},
          {"balioa" : "MOTIB", "izena" : "Motibatiboa"},
          {"balioa" : "ONDO", "izena" : "Ondoriozkoa"},
          {"balioa" : "ZHG", "izena" : "Zehar-galdera"}
        ]},
        {"Kasua" : [
          {"balioa" : "ABL", "izena" : "Ablatiboa"},
          {"balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},

```

```

{"balioa" : "ABS", "izena" : "Absolutiboa"},
{"balioa" : "ABZ", "izena" : "Adlatibo bide zuzenekoa"},
{"balioa" : "ALA", "izena" : "Adlatiboa"},
{"balioa" : "BNK", "izena" : "Bankaria"},
{"balioa" : "DAT", "izena" : "Datiboa"},
{"balioa" : "DES", "izena" : "Destinatiboa"},
{"balioa" : "DESK", "izena" : "Deskribatzailea"},
{"balioa" : "ERG", "izena" : "Ergatiboa"},
{"balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa"},
{"balioa" : "GEN", "izena" : "Genitibo edutezkoa"},
{"balioa" : "INE", "izena" : "Inesiboa"},
{"balioa" : "INS", "izena" : "Instrumentala"},
{"balioa" : "MOT", "izena" : "Motibatiboa"},
{"balioa" : "PAR", "izena" : "Partitiboa"},
{"balioa" : "PRO", "izena" : "Prolatiboa"},
{"balioa" : "SOZ", "izena" : "Soziatiboa"}

```

```

]},

```

```

{"LaguntzaileMota" : [

```

```

    {"balioa" : "DA", "izena" : "Da"},
    {"balioa" : "DU", "izena" : "Du"},
    {"balioa" : "DA-DU", "izena" : "Da-du"},
    {"balioa" : "DIO", "izena" : "Dio"},
    {"balioa" : "ZAI", "izena" : "Zai"}

```

```

]},

```

```

{"ModuaDenbora" : [

```

```

    {"balioa" : "A1", "izena" : "Indikatibozko orainaldia (naiz)"},
    {"balioa" : "A2", "izena" : "Indikatibozko geroaldi arkaikoa (nintzateke)"},
    {"balioa" : "A3", "izena" : "Subjuntibozko orainaldia (nadin)"},
    {"balioa" : "A4", "izena" : "Subjuntibozko baldintza orainaldia (banadi)"},
    {"balioa" : "A5", "izena" : "Ahalezko orainaldia (naiteke)"},
    {"balioa" : "B1", "izena" : "Indikatibozko lehenaldia (nintzen)"},
    {"balioa" : "B2", "izena" : "Indikatibozko baldintza, ondorioa/orain-gero

```

```

(nintzateke)"},

```

```

    {"balioa" : "B3", "izena" : "Indikatibozko baldintza, ondorioa/lehen (nintzatekeen)"},
    {"balioa" : "B4", "izena" : "Indikatibozko baldintza, aurrekoa (banintz)"},
    {"balioa" : "B5A", "izena" : "Subjuntibozko alegiazkoa (ledin)"},
    {"balioa" : "B5B", "izena" : "Subjuntibozko lehenaldia (zedin)"},
    {"balioa" : "B6", "izena" : "Subjuntibozko baldintza, lehenaldia (banendi)"},
    {"balioa" : "B7", "izena" : "Ahalezko lehenaldia (nintzateke)"},
    {"balioa" : "B8", "izena" : "Ahalezko lehenaldi urruna (nintzatekeen)"},
    {"balioa" : "C", "izena" : "Inperatibozko orainaldia (hadi)}

```

```

]},

```

```

{"Numeroa" : [

```

```

    {"balioa" : "S", "izena" : "Singularra"},
    {"balioa" : "P", "izena" : "Plurala"},
    {"balioa" : "PH", "izena" : "Plural hurbila"}

```

```

    }}
  ]
},
{ "balioa" : "ADJ",
  "izena" : "Adjektiboa",
  "ezaugIzenak" : [ "Azpikategoriak", "IzenarenAurretik", "Kasua", "Mugatasuna", "Numeroa" ],
  "Ezaugarriak" : [
    { "Azpikategoriak" : [
      { "balioa" : "ARR", "izena" : "Adj. arrunta" },
      { "balioa" : "GAL", "izena" : "Adj. galdetzailea" }
    ] },
    { "IzenarenAurretik" : [
      { "balioa" : "BAI", "izena" : "Bai" },
      { "balioa" : "EZ", "izena" : "Ez" }
    ] },
    { "Kasua" : [
      { "balioa" : "ABL", "izena" : "Ablatiboa" },
      { "balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa" },
      { "balioa" : "ABS", "izena" : "Absolutiboa" },
      { "balioa" : "ABZ", "izena" : "Adlatibo bide zuzenekoa" },
      { "balioa" : "ALA", "izena" : "Adlatiboa" },
      { "balioa" : "BNK", "izena" : "Bankaria" },
      { "balioa" : "DAT", "izena" : "Datiboa" },
      { "balioa" : "DES", "izena" : "Destinatiboa" },
      { "balioa" : "DESK", "izena" : "Deskribatzailea" },
      { "balioa" : "ERG", "izena" : "Ergatiboa" },
      { "balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa" },
      { "balioa" : "GEN", "izena" : "Genitibo edutezkoa" },
      { "balioa" : "INE", "izena" : "Inesiboa" },
      { "balioa" : "INS", "izena" : "Instrumentala" },
      { "balioa" : "MOT", "izena" : "Motibatiboa" },
      { "balioa" : "PAR", "izena" : "Partitiboa" },
      { "balioa" : "PRO", "izena" : "Prolatiboa" },
      { "balioa" : "SOZ", "izena" : "Soziatiboa" }
    ] },
    { "Mugatasuna" : [
      { "balioa" : "M", "izena" : "Mugatua" },
      { "balioa" : "MG", "izena" : "Mugagabea" }
    ] },
    { "Numeroa" : [
      { "balioa" : "S", "izena" : "Singularra" },
      { "balioa" : "P", "izena" : "Plurala" },
      { "balioa" : "PH", "izena" : "Plural hurbila" }
    ] }
  ]
}

```

```

{ "balioa" : "ADB",
  "izena" : "Adberbioa",
  "ezaugIzenak" : [ "Azpikategoriak", "AdberbioMota", "Kasua", "GraduMaila", "Mugatasuna", "Numeroa" ],
  "Ezaugarriak" : [
    { "Azpikategoriak" : [
      { "balioa" : "ARR", "izena" : "Adb. arrunta"},
      { "balioa" : "GAL", "izena" : "Adb. galdetzailea"}
    ]},
    { "AdberbioMota" : [
      { "balioa" : "DENB", "izena" : "Denbora"},
      { "balioa" : "MOD", "izena" : "Modua"},
      { "balioa" : "GRAD", "izena" : "Gradua"},
      { "balioa" : "LEK", "izena" : "Lekua"}
    ]},
    { "Kasua" : [
      { "balioa" : "ABL", "izena" : "Ablatiboa"},
      { "balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},
      { "balioa" : "ABS", "izena" : "Absolutiboa"},
      { "balioa" : "ABZ", "izena" : "Adlatibo bide zuzeneko"},
      { "balioa" : "ALA", "izena" : "Adlatiboa"},
      { "balioa" : "BNK", "izena" : "Bankaria"},
      { "balioa" : "DAT", "izena" : "Datiboa"},
      { "balioa" : "DES", "izena" : "Destinatiboa"},
      { "balioa" : "DESK", "izena" : "Deskribatzailea"},
      { "balioa" : "ERG", "izena" : "Ergatiboa"},
      { "balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa"},
      { "balioa" : "GEN", "izena" : "Genitibo edutezkoa"},
      { "balioa" : "INE", "izena" : "Inesiboa"},
      { "balioa" : "INS", "izena" : "Instrumentala"},
      { "balioa" : "MOT", "izena" : "Motibatiboa"},
      { "balioa" : "PAR", "izena" : "Partitiboa"},
      { "balioa" : "PRO", "izena" : "Prolatiboa"},
      { "balioa" : "SOZ", "izena" : "Soziatiboa"}
    ]},
    { "GraduMaila" : [
      { "balioa" : "KONP", "izena" : "Konparatiboa"},
      { "balioa" : "GEHI", "izena" : "Gehiegizkoa"},
      { "balioa" : "SUP", "izena" : "Superlatiboa"},
      { "balioa" : "IND", "izena" : "Indargarria"}
    ]},
    { "Mugatasuna" : [
      { "balioa" : "M", "izena" : "Mugatua"},
      { "balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    { "Numeroa" : [
      { "balioa" : "S", "izena" : "Singularra"},

```

```

        {"balioa" : "P", "izena" : "Plurala"},
        {"balioa" : "PH", "izena" : "Plural hurbila"}
    ]}

]
},
{"balioa" : "IOR", "izena" : "Izen ordaina",
"ezaugIzenak" : [ "Azpikategoriak", "Kasua", "Mugatasuna", "NumeroaMugatasuna", "Numeroa",
    "Pertsona" ],
"Ezaugarriak" : [
    {"Azpikategoriak" : [
        {"balioa" : "PERARR", "izena" : "Pertsonal arrunta"},
        {"balioa" : "IZGMGB", "izena" : "Zehaztugabe mugagabea"},
        {"balioa" : "PERIND", "izena" : "Pertsonal indartua"},
        {"balioa" : "ELK", "izena" : "Elkarkaria"},
        {"balioa" : "IZGAL", "izena" : "Galdetzailea"}
    ]},
    {"Kasua" : [
        {"balioa" : "ABL", "izena" : "Ablatiboa"},
        {"balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},
        {"balioa" : "ABS", "izena" : "Absolutiboa"},
        {"balioa" : "ABZ", "izena" : "Adlatibo bide zuzeneko"},
        {"balioa" : "ALA", "izena" : "Adlatiboa"},
        {"balioa" : "BNK", "izena" : "Bankaria"},
        {"balioa" : "DAT", "izena" : "Datiboa"},
        {"balioa" : "DES", "izena" : "Destinatiboa"},
        {"balioa" : "DESK", "izena" : "Deskribatzailea"},
        {"balioa" : "ERG", "izena" : "Ergatiboa"},
        {"balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa"},
        {"balioa" : "GEN", "izena" : "Genitibo edutezkoa"},
        {"balioa" : "INE", "izena" : "Inesiboa"},
        {"balioa" : "INS", "izena" : "Instrumentala"},
        {"balioa" : "MOT", "izena" : "Motibatiboa"},
        {"balioa" : "PAR", "izena" : "Partitiboa"},
        {"balioa" : "PRO", "izena" : "Prolatiboa"},
        {"balioa" : "SOZ", "izena" : "Soziatiboa"}
    ]},
    {"Mugatasuna" : [
        {"balioa" : "M", "izena" : "Mugatua"},
        {"balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    {"NumeroaMugatasuna" : [
        {"balioa" : "S", "izena" : "Singularra"},
        {"balioa" : "P", "izena" : "Plurala"},
        {"balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    {"Numeroa" : [

```

```

        {"balioa" : "S", "izena" : "Singularra"},
        {"balioa" : "P", "izena" : "Plurala"},
        {"balioa" : "PH", "izena" : "Plural hurbila"}
    ]},
    {"Pertsona" : [
        {"balioa" : "NI", "izena" : "Ni"},
        {"balioa" : "HI", "izena" : "Hi"},
        {"balioa" : "HURA", "izena" : "Hura"},
        {"balioa" : "GU", "izena" : "Gu"},
        {"balioa" : "ZU", "izena" : "Zu"},
        {"balioa" : "ZUEK", "izena" : "Zuek"},
        {"balioa" : "HAIEK", "izena" : "Haiek"}
    ]}
]
},
{"balioa" : "ADL", "izena" : "Aditz laguntzailea",
"ezaugIzenak" : ["Erlazioa", "Hitanoa", "Kasua", "ModuaDenbora", "Mugatasuna", "Nor", "Nori", "Nork"],
"Ezaugarriak" : [
    {"Erlazioa" : [
        {"balioa" : "AURK", "izena" : "Aurkaritzakoa"},
        {"balioa" : "BALD", "izena" : "Baldintzazkoa"},
        {"balioa" : "DENB", "izena" : "Denborazkoa"},
        {"balioa" : "EMEN", "izena" : "Emendiozkoa"},
        {"balioa" : "ERLT", "izena" : "Erlatibozkoa"},
        {"balioa" : "ESPL", "izena" : "Esplikatiboa"},
        {"balioa" : "HAUT", "izena" : "Hautakaria"},
        {"balioa" : "HELB", "izena" : "Helburuzkoa"},
        {"balioa" : "KAUS", "izena" : "Kausazkoa"},
        {"balioa" : "KONPL", "izena" : "Konpletiboa"},
        {"balioa" : "KONT", "izena" : "Kontzesiboa"},
        {"balioa" : "MOD", "izena" : "Moduzkoa"},
        {"balioa" : "MOD/DENB", "izena" : "Moduzkoa/Denborazkoa"},
        {"balioa" : "MOS", "izena" : "Menderagailu-osagaia"},
        {"balioa" : "MOTIB", "izena" : "Motibatiboa"},
        {"balioa" : "ONDO", "izena" : "Ondoriozkoa"},
        {"balioa" : "ZHG", "izena" : "Zehar-galdera"}
    ]},
    {"Hitanoa" : [
        {"balioa" : "TO", "izena" : "To"},
        {"balioa" : "NO", "izena" : "No"}
    ]},
    {"Kasua" : [
        {"balioa" : "ABL", "izena" : "Ablatiboa"},
        {"balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},
        {"balioa" : "ABS", "izena" : "Absolutiboa"},
        {"balioa" : "ABZ", "izena" : "Adlatibo bide zuzeneko"},

```



```

        {"balioa": "ALA", "izena": "Adlatiboa"},
        {"balioa": "BNK", "izena": "Bankaria"},
        {"balioa": "DAT", "izena": "Datiboa"},
        {"balioa": "DES", "izena": "Destinatiboa"},
        {"balioa": "DESK", "izena": "Deskribatzailea"},
        {"balioa": "ERG", "izena": "Ergatiboa"},
        {"balioa": "GEL", "izena": "Genitibo leku-denborazkoa"},
        {"balioa": "GEN", "izena": "Genitibo edutezkoa"},
        {"balioa": "INE", "izena": "Inesiboa"},
        {"balioa": "INS", "izena": "Instrumentala"},
        {"balioa": "MOT", "izena": "Motibatiboa"},
        {"balioa": "PAR", "izena": "Partitiboa"},
        {"balioa": "PRO", "izena": "Prolatiboa"},
        {"balioa": "SOZ", "izena": "Soziatiboa"}
    ]},
    {"ModuaDenbora": [
        {"balioa": "A1", "izena": "Indikatibozko orainaldia (naiz)"},
        {"balioa": "A2", "izena": "Indikatibozko geroaldi arkaikoa (nintzateke)"},
        {"balioa": "A3", "izena": "Subjuntibozko orainaldia (nadin)"},
        {"balioa": "A4", "izena": "Subjuntibozko baldintza orainaldia (banadi)"},
        {"balioa": "A5", "izena": "Ahalezko orainaldia (naiteke)"},
        {"balioa": "B1", "izena": "Indikatibozko lehenaldia (nintzen)"},
        {"balioa": "B2", "izena": "Indikatibozko baldintza, ondorioa/orain-gero  

            (nintzateke)"},
        {"balioa": "B3", "izena": "Indikatibozko baldintza, ondorioa/lehen (nintzatekeen)"},
        {"balioa": "B4", "izena": "Indikatibozko baldintza, aurrekoa (banintz)"},
        {"balioa": "B5A", "izena": "Subjuntibozko alegiazkoa (ledin)"},
        {"balioa": "B5B", "izena": "Subjuntibozko lehenaldia (zedin)"},
        {"balioa": "B6", "izena": "Subjuntibozko baldintza, lehenaldia (banendi)"},
        {"balioa": "B7", "izena": "Ahalezko lehenaldia (ninteke)"},
        {"balioa": "B8", "izena": "Ahalezko lehenaldi urruna (nintzekeen)"},
        {"balioa": "C", "izena": "Inperatibozko orainaldia (hadi)}
    ]},
    {"Mugatasuna": [
        {"balioa": "M", "izena": "Mugatua"},
        {"balioa": "MG", "izena": "Mugagabea"}
    ]},
    {"Nor": [
        {"balioa": "NI", "izena": "Ni"},
        {"balioa": "HI", "izena": "Hi"},
        {"balioa": "HURA", "izena": "Hura"},
        {"balioa": "GU", "izena": "Gu"},
        {"balioa": "ZU", "izena": "Zu"},
        {"balioa": "ZUEK", "izena": "Zuek"},
        {"balioa": "HAIEK", "izena": "Haiek"}
    ]},
    ]},

```

```

{"Nori" : [
    {"balioa" : "NIRI", "izena" : "Niri"},
    {"balioa" : "HIRI-TO", "izena" : "Hiri (toka)"},
    {"balioa" : "HIRI-NO", "izena" : "Hiri (noka)"},
    {"balioa" : "HARI", "izena" : "Hari"},
    {"balioa" : "GURI", "izena" : "Guri"},
    {"balioa" : "ZURI", "izena" : "Zuri"},
    {"balioa" : "ZUEI", "izena" : "Zuei"},
    {"balioa" : "HAIEI", "izena" : "Haiei"}
]},
{"Nork" : [
    {"balioa" : "NIK", "izena" : "Nik"},
    {"balioa" : "HIK", "izena" : "Hik"},
    {"balioa" : "HIK-TO", "izena" : "Hik (toka)"},
    {"balioa" : "HIK-NO", "izena" : "Hik (noka)"},
    {"balioa" : "HARK", "izena" : "Hark"},
    {"balioa" : "GUK", "izena" : "Guk"},
    {"balioa" : "ZUK", "izena" : "Zuk"},
    {"balioa" : "ZUEK", "izena" : "Zuek"},
    {"balioa" : "HAIEK", "izena" : "Haiek"}
]}
]
},
{"balioa" : "ADT", "izena" : "Aditz trinkoa",
"ezaugIzenak" : ["Aspektua", "Erlazioa", "Hitanoa", "Kasua", "ModuaDenbora", "Mugatasuna", "Nor",
                  "Nori", "Nork"],
"Ezaugarriak" : [
    {"Aspektua" : [
        {"balioa" : "GERO", "izena" : "Etorkizuna"},
        {"balioa" : "BURU", "izena" : "Burutua"},
        {"balioa" : "PNT", "izena" : "Puntukaria"},
        {"balioa" : "EZBU", "izena" : "Burutugabea"}
    ]},
    {"Erlazioa" : [
        {"balioa" : "AURK", "izena" : "Aurkaritzakoa"},
        {"balioa" : "BALD", "izena" : "Baldintzazkoa"},
        {"balioa" : "DENB", "izena" : "Denborazkoa"},
        {"balioa" : "EMEN", "izena" : "Emendiozkoa"},
        {"balioa" : "ERLT", "izena" : "Erlatibozkoa"},
        {"balioa" : "ESPL", "izena" : "Esplikatiboa"},
        {"balioa" : "HAUT", "izena" : "Hautakaria"},
        {"balioa" : "HELB", "izena" : "Helburuzkoa"},
        {"balioa" : "KAUS", "izena" : "Kausazkoa"},
        {"balioa" : "KONPL", "izena" : "Konpletiboa"},
        {"balioa" : "KONT", "izena" : "Kontzesiboa"},
        {"balioa" : "MOD", "izena" : "Moduzkoa"}
    ]}
    ]
}

```

```

        {"balioa" : "MOD/DENB", "izena" : "Moduzkoa/Denborazkoa"},
        {"balioa" : "MOS", "izena" : "Menderagailu-osagaia"},
        {"balioa" : "MOTIB", "izena" : "Motibatiboa"},
        {"balioa" : "ONDO", "izena" : "Ondoriozkoa"},
        {"balioa" : "ZHG", "izena" : "Zehar-galdera"}
    ]},
    {"Hitanoa" : [
        {"balioa" : "TO", "izena" : "To"},
        {"balioa" : "NO", "izena" : "No"}
    ]},
    {"Kasua" : [
        {"balioa" : "ABL", "izena" : "Ablatiboa"},
        {"balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},
        {"balioa" : "ABS", "izena" : "Absolutiboa"},
        {"balioa" : "ABZ", "izena" : "Adlatibo bide zuzenekoa"},
        {"balioa" : "ALA", "izena" : "Adlatiboa"},
        {"balioa" : "BNK", "izena" : "Bankaria"},
        {"balioa" : "DAT", "izena" : "Datiboa"},
        {"balioa" : "DES", "izena" : "Destinatiboa"},
        {"balioa" : "DESK", "izena" : "Deskribatzailea"},
        {"balioa" : "ERG", "izena" : "Ergatiboa"},
        {"balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa"},
        {"balioa" : "GEN", "izena" : "Genitibo edutezkoa"},
        {"balioa" : "INE", "izena" : "Inesiboa"},
        {"balioa" : "INS", "izena" : "Instrumentala"},
        {"balioa" : "MOT", "izena" : "Motibatiboa"},
        {"balioa" : "PAR", "izena" : "Partitiboa"},
        {"balioa" : "PRO", "izena" : "Prolatiboa"},
        {"balioa" : "SOZ", "izena" : "Soziatiboa"}
    ]},
    {"ModuaDenbora" : [
        {"balioa" : "A1", "izena" : "Indikatibozko orainaldia (naiz)"},
        {"balioa" : "A2", "izena" : "Indikatibozko geroaldi arkaikoa (nintzateke)"},
        {"balioa" : "A3", "izena" : "Subjuntibozko orainaldia (nadin)"},
        {"balioa" : "A4", "izena" : "Subjuntibozko baldintza orainaldia (banadi)"},
        {"balioa" : "A5", "izena" : "Ahalezko orainaldia (naiteke)"},
        {"balioa" : "B1", "izena" : "Indikatibozko lehenaldia (nintzen)"},
        {"balioa" : "B2", "izena" : "Indikatibozko baldintza, ondorioa/orain-gero (nintzateke)"},
        {"balioa" : "B3", "izena" : "Indikatibozko baldintza, ondorioa/lehen (nintzatekeen)"},
        {"balioa" : "B4", "izena" : "Indikatibozko baldintza, aurrekoa (banintz)"},
        {"balioa" : "B5A", "izena" : "Subjuntibozko alegiazkoa (ledin)"},
        {"balioa" : "B5B", "izena" : "Subjuntibozko lehenaldia (zedin)"},
        {"balioa" : "B6", "izena" : "Subjuntibozko baldintza, lehenaldia (banendi)"},
        {"balioa" : "B7", "izena" : "Ahalezko lehenaldia (ninteke)"},
        {"balioa" : "B8", "izena" : "Ahalezko lehenaldi urruna (nintzeen)"}
    ]}

```

```

        {"balioa" : "C", "izena" : "Inperatibozko orainaldia (hadi)"}
    ]},
    {"Mugatasuna" : [
        {"balioa" : "M", "izena" : "Mugatua"},
        {"balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    {"Nor" : [
        {"balioa" : "NI", "izena" : "Ni"},
        {"balioa" : "HI", "izena" : "Hi"},
        {"balioa" : "HURA", "izena" : "Hura"},
        {"balioa" : "GU", "izena" : "Gu"},
        {"balioa" : "ZU", "izena" : "Zu"},
        {"balioa" : "ZUEK", "izena" : "Zuek"},
        {"balioa" : "HAIEK", "izena" : "Haiek"}
    ]},
    {"Nori" : [
        {"balioa" : "NIRI", "izena" : "Niri"},
        {"balioa" : "HIRI-TO", "izena" : "Hiri (toka)"},
        {"balioa" : "HIRI-NO", "izena" : "Hiri (noka)"},
        {"balioa" : "HARI", "izena" : "Hari"},
        {"balioa" : "GURI", "izena" : "Guri"},
        {"balioa" : "ZURI", "izena" : "Zuri"},
        {"balioa" : "ZUEI", "izena" : "Zuei"},
        {"balioa" : "HAIEI", "izena" : "Haiei"}
    ]},
    {"Nork" : [
        {"balioa" : "NIK", "izena" : "Nik"},
        {"balioa" : "HIK", "izena" : "Hik"},
        {"balioa" : "HIK-TO", "izena" : "Hik (toka)"},
        {"balioa" : "HIK-NO", "izena" : "Hik (noka)"},
        {"balioa" : "HARK", "izena" : "Hark"},
        {"balioa" : "GUK", "izena" : "Guk"},
        {"balioa" : "ZUK", "izena" : "Zuk"},
        {"balioa" : "ZUEK", "izena" : "Zuek"},
        {"balioa" : "HAIEK", "izena" : "Haiek"}
    ]}
]
},
{"balioa" : "DET", "izena" : "Determinatzailea",
"ezaugarriak" : [ "Azpikategoriak", "Kasua", "Mugatasuna", "NumeroaMugatasuna", "Numeroa"],
"Ezaugarriak" : [
    {"Azpikategoriak" : [
        {"balioa" : "DZH", "izena" : "Zenb. zehaztua"},
        {"balioa" : "ERKARR", "izena" : "Erak. arrunta"},
        {"balioa" : "DZG", "izena" : "Zenb. zehaztugabea"},
        {"balioa" : "ERKIND", "izena" : "Erak. indartua"},
    ]}
]

```

```

        {"balioa" : "ORD", "izena" : "Zenb. zehaztu ordinala"},
        {"balioa" : "ORO", "izena" : "Zenb. orokorra"},
        {"balioa" : "NOLGAL", "izena" : "Nolakotzaile galdetzailea"},
        {"balioa" : "NOLARR", "izena" : "Nolakotzaile arrunta"},
        {"balioa" : "BAN", "izena" : "Zenb. zehaztu banatzailea"}
    ]},
    {"Kasua" : [
        {"balioa" : "ABL", "izena" : "Ablatiboa"},
        {"balioa" : "ABU", "izena" : "Adlatibo bukatuzkoa"},
        {"balioa" : "ABS", "izena" : "Absolutiboa"},
        {"balioa" : "ABZ", "izena" : "Adlatibo bide zuzenekoa"},
        {"balioa" : "ALA", "izena" : "Adlatiboa"},
        {"balioa" : "BNK", "izena" : "Bankaria"},
        {"balioa" : "DAT", "izena" : "Datiboa"},
        {"balioa" : "DES", "izena" : "Destinatiboa"},
        {"balioa" : "DESK", "izena" : "Deskribatzailea"},
        {"balioa" : "ERG", "izena" : "Ergatiboa"},
        {"balioa" : "GEL", "izena" : "Genitibo leku-denborazkoa"},
        {"balioa" : "GEN", "izena" : "Genitibo edutezkoa"},
        {"balioa" : "INE", "izena" : "Inesiboa"},
        {"balioa" : "INS", "izena" : "Instrumentala"},
        {"balioa" : "MOT", "izena" : "Motibatiboa"},
        {"balioa" : "PAR", "izena" : "Partitiboa"},
        {"balioa" : "PRO", "izena" : "Prolatiboa"},
        {"balioa" : "SOZ", "izena" : "Soziatiboa"}
    ]},
    {"Mugatasuna" : [
        {"balioa" : "M", "izena" : "Mugatua"},
        {"balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    {"NumeroaMugatasuna" : [
        {"balioa" : "S", "izena" : "Singularra"},
        {"balioa" : "P", "izena" : "Plurala"},
        {"balioa" : "MG", "izena" : "Mugagabea"}
    ]},
    {"Numeroa" : [
        {"balioa" : "S", "izena" : "Singularra"},
        {"balioa" : "P", "izena" : "Plurala"},
        {"balioa" : "PH", "izena" : "Plural hurbila"}
    ]}
]
},
{"balioa" : "ITJ", "izena" : "Interjekzioa",
"ezaugIzenak" : [ "Azpikategoriak", "Kasua", "Mugatasuna" ],
"Ezaugarriak" : [
    {"Azpikategoriak" : [

```

```

    ]},
    {"Kasua": [
        {"balioa": "ABL", "izena": "Ablatiboa"},
        {"balioa": "ABU", "izena": "Adlatibo bukatuzkoa"},
        {"balioa": "ABS", "izena": "Absolutiboa"},
        {"balioa": "ABZ", "izena": "Adlatibo bide zuzeneko"},
        {"balioa": "ALA", "izena": "Adlatiboa"},
        {"balioa": "BNK", "izena": "Bankaria"},
        {"balioa": "DAT", "izena": "Datiboa"},
        {"balioa": "DES", "izena": "Destinatiboa"},
        {"balioa": "DESK", "izena": "Deskribatzailea"},
        {"balioa": "ERG", "izena": "Ergatiboa"},
        {"balioa": "GEL", "izena": "Genitibo leku-denborazkoa"},
        {"balioa": "GEN", "izena": "Genitibo edutezkoa"},
        {"balioa": "INE", "izena": "Inesiboa"},
        {"balioa": "INS", "izena": "Instrumentala"},
        {"balioa": "MOT", "izena": "Motibatiboa"},
        {"balioa": "PAR", "izena": "Partitiboa"},
        {"balioa": "PRO", "izena": "Prolatiboa"},
        {"balioa": "SOZ", "izena": "Soziatiboa"}
    ]},
    {"Mugatasuna": [
        {"balioa": "M", "izena": "Mugatua"},
        {"balioa": "MG", "izena": "Mugagabea"}
    ]}
]
},
{"balioa": "LOT", "izena": "Loturazkoa",
"ezaugarriak": ["Azpikategoriak", "Erlazioa", "KlausulaMuga"],
"Ezaugarriak": [
    {"Azpikategoriak": [
        {"balioa": "LOK", "izena": "Lokailua"},
        {"balioa": "JNT", "izena": "Juntagailua"},
        {"balioa": "MEN", "izena": "Menderagailua"}
    ]},
    {"Erlazioa": [
        {"balioa": "AURK", "izena": "Aurkaritzakoa"},
        {"balioa": "BALD", "izena": "Baldintzazkoa"},
        {"balioa": "DENB", "izena": "Denborazkoa"},
        {"balioa": "EMEN", "izena": "Emendiozkoa"},
        {"balioa": "ERLT", "izena": "Erlatibozkoa"},
        {"balioa": "ESPL", "izena": "Esplikatiboa"},
        {"balioa": "HAUT", "izena": "Hautakaria"},
        {"balioa": "HELB", "izena": "Helburuzkoa"},
        {"balioa": "KAUS", "izena": "Kausazkoa"},
        {"balioa": "KONPL", "izena": "Konpletiboa"}
    ]}
    ]}

```

```

        {"balioa" : "KONT", "izena" : "Kontzesiboa"},
        {"balioa" : "MOD", "izena" : "Moduzkoa"},
        {"balioa" : "MOD/DENB", "izena" : "Moduzkoa/Denborazkoa"},
        {"balioa" : "MOS", "izena" : "Menderagailu-osagaia"},
        {"balioa" : "MOTIB", "izena" : "Motibatiboa"},
        {"balioa" : "ONDO", "izena" : "Ondoriozkoa"},
        {"balioa" : "ZHG", "izena" : "Zehar-galdera"}
    ]},
    {"KlausulaMuga" : [
        {"balioa" : "HAS", "izena" : "Klausula-hasiera"},
        {"balioa" : "AM", "izena" : "Klausula-amaiera"},
        {"balioa" : "HA", "izena" : "Klausula -hasiera edo -amaiera"}
    ]}
]
},
{"balioa" : "PRT", "izena" : "Partikula",
"ezaugIzenak" : [ "Erlazioa", "Modalitatea" ],
"Ezaugarriak" : [
    {"Erlazioa" : [
        {"balioa" : "AURK", "izena" : "Aurkaritzakoa"},
        {"balioa" : "BALD", "izena" : "Baldintzazkoa"},
        {"balioa" : "DENB", "izena" : "Denborazkoa"},
        {"balioa" : "EMEN", "izena" : "Emendiozkoa"},
        {"balioa" : "ERLT", "izena" : "Erlatibozkoa"},
        {"balioa" : "ESPL", "izena" : "Esplikatiboa"},
        {"balioa" : "HAUT", "izena" : "Hautakaria"},
        {"balioa" : "HELB", "izena" : "Helburuzkoa"},
        {"balioa" : "KAUS", "izena" : "Kausazkoa"},
        {"balioa" : "KONPL", "izena" : "Konpletiboa"},
        {"balioa" : "KONT", "izena" : "Kontzesiboa"},
        {"balioa" : "MOD", "izena" : "Moduzkoa"},
        {"balioa" : "MOD/DENB", "izena" : "Moduzkoa/Denborazkoa"},
        {"balioa" : "MOS", "izena" : "Menderagailu-osagaia"},
        {"balioa" : "MOTIB", "izena" : "Motibatiboa"},
        {"balioa" : "ONDO", "izena" : "Ondoriozkoa"},
        {"balioa" : "ZHG", "izena" : "Zehar-galdera"}
    ]},
    {"Modalitatea" : [
        {"balioa" : "ZIU", "izena" : "Ziurtasunezkoa"},
        {"balioa" : "EGI", "izena" : "Egiatasunezkoa"}
    ]}
]
},
{"balioa" : "BST", "izena" : "Bestelakoa",
"ezaugIzenak" : [ "Azpikategoriak", "Kasua", "Mugatasuna" ],
"Ezaugarriak" : [

```

```

{"Azpikategoriak": [
    {"balioa": "ARR", "izena": "Bestelako arrunta"}
]},
{"Kasua": [
    {"balioa": "ABL", "izena": "Ablatiboa"},
    {"balioa": "ABU", "izena": "Adlatibo bukatuzkoa"},
    {"balioa": "ABS", "izena": "Absolutiboa"},
    {"balioa": "ABZ", "izena": "Adlatibo bide zuzeneko"},
    {"balioa": "ALA", "izena": "Adlatiboa"},
    {"balioa": "BNK", "izena": "Banakaria"},
    {"balioa": "DAT", "izena": "Datiboa"},
    {"balioa": "DES", "izena": "Destinatiboa"},
    {"balioa": "DESK", "izena": "Deskribatzailea"},
    {"balioa": "ERG", "izena": "Ergatiboa"},
    {"balioa": "GEL", "izena": "Genitibo leku-denborazkoa"},
    {"balioa": "GEN", "izena": "Genitibo edutezkoa"},
    {"balioa": "INE", "izena": "Inesiboa"},
    {"balioa": "INS", "izena": "Instrumentala"},
    {"balioa": "MOT", "izena": "Motibatiboa"},
    {"balioa": "PAR", "izena": "Partitiboa"},
    {"balioa": "PRO", "izena": "Prolatiboa"},
    {"balioa": "SOZ", "izena": "Soziatiboa"}
]},
{"Mugatasuna": [
    {"balioa": "M", "izena": "Mugatua"},
    {"balioa": "MG", "izena": "Mugagabea"}
]}
]
}

```


2. ERANSKINA

Bilera-aktak

0.Bilera-akta

Data: 2014-10-10

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 11:00

Amaiera ordua: 11:45

Hitz egindakoa:

- Proiektuaren helburua labur azaldu.
- Nire betebeharra zein den zehaztu.
- Komunikaziorako erabiliko dugun mekanismoa zehaztu: ostiralero 11tan niretzako erreserbatu dute ordua, baina aurrez baieztatu beharko dut bilera.

Egin beharrekoa:

- *Elasticsearchi* buruz informazio pixka bat bilatu.

1.Bilera-akta

Data: 2014-10-23

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 16:00

Amaiera ordua: 16:45

Hitz egindakoa:

- Filmak indizea nola sortu nuen erakutsi diet (Google Driven daukat azalduta pausoz pauso), KOPF online-plataformarekin.

Egin beharrekoa:

- Kibana instalatu. *Elasticsearch*en webgune ofizialean interfaze horrekin egiten dituzte gauza guztiak eta informazio gehiago dago.
- DEMO bat egin testuak indizea sortuz. Testu batek bi propietate izango ditu: izenburua eta testua bera.

2.Bilera-akta

Data: 2014-11-06

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 16:00

Amaiera ordua: 16:30

Hitz egindakoa:

- Kibana instalatu dudala erakutsi diet eta testuak gordetzeko indizea ere bai.

Egin beharrekoa:

- Hartutako erabakia: Bilerak ostiral goizetik ostegun arratsaldeetara pasa ditugu.
- CURL, JSON-ekin... familiarizatu. Soltura hartu.
- Stemming-ak nola egiten diren begiratu. (Stemming: 'tables' hitza edukita, 'table' jartzea. Hau da, hitz guztiak *izen* forman jartzea.)
- Leihoena nola egiten den begiratu. (Leioa: hitz baten ondoan zein beste hitzak joan behar duen. Gerta daiteke bost leihoko hitza izatea)
- Testu normalizatuak nola egiten diren begiratu.

3.Bilera-akta

Data: 2014-11-27

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 16:00

Amaiera ordua: 16:45

Hitz egindakoa:

- [Stemming](#) eta [leihoekin](#) (nGram) egindakoa erakutsi.
- Stemming-a eginda dago, beraz, horretan ez naiz sartuko.

Egin beharrekoa:

- Field-ak sortu

testua	etxetik	zeturten	goizean
stemming (eginda dago)	etxe	etorri	goiz
mota	IZE	ADI	IZE

- Leihoak (slop) begiratu.
- Bilaketak egiten saiatu horrekin.

4.Bilera-akta

Data: 2014-12-11

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 16:00

Amaiera ordua: 16:30

Hitz egin beharrekoa:

- *testuak2* indize berria sortu dudala erakutsi.
- Stemmerrarena ondo doala erakutsi (*MultifieldBilaketak.json* eta *Testuak(MultifieldBilaketak).json* fitxategiak).
- Leihoen arazoa azaldu: Bidaltzen dugun aginduak ez du guk nahi duguna itzultzen.

Egin beharrekoa:

- Leihoen arazo hori, momentuz, horrela utzi. Aurrerago begiratuko dugu.
- Kategoriaren eskema aldatu. Ez erabili 'field' horretarako.
- Leihoe kin bilaketak hitzen bidez egin ordeztu, eremuekin egiten saiatu.

5.Bilera-akta

Data: 2015-02-18

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:00

Hitz egin beharrekoa:

- Leihoen bidezko bilaketak eremuen bidez egiten saiatu naiz, baina ez dut guztiz lortu. *Testuak*(*MultifieldBilaketak*).json fitxategian erakutsi adibideak.
 - “exists” eremuaren laguntzaz.
 - Espresio erregularren laguntzaz.

Zer nahi dugu:

Dokumentua: “Etxetik nentorren”

Hitz bakoitzari zenbait informazio erantsiko diogu:

“Etxetik [“etxe”, IZE, IZARRUNTA, ABLATIBOA] nentorren [“etorri”, ADI, ADTrinkoa, LEHENALDIA, ...]”

[lema, kat, azp, kasua]

1) Galdera arruntak:

- Bilatu kat(“IZE”) duten dokumentuak
 - “Etxetik nentorren”
- Bilatu kat(“ADTrinkoa”) duten dokumentuak
 - “Etxetik nentorren”
- Bilatu lema(“etorri”)
- Bilatu “Etxetik”

2) Leihoak

- Bilatu kat(“IZE”) + kat(“ADTrinkoa”) duten dokumentuak
 - “Etxetik nentorren”
- Bilatu “Etxetik” + lema(“etorri”)

3) Anbiguotasuna

- analisi bat baino gehiago esleitu ahal izatea hitz bati
 - “txakurrak” ([IZE, ABS, PLU], [IZE, ERG, SING])

6.Bilera-akta

Data: 2015-03-04

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:50

Hitz egin beharrekoa:

- Aurreko bilerako bigarren puntua (leihoena) ez dut lortu.

Egin beharrekoa:

- Eskema aldatuz, leihoena probatu:
 - Esaldiak
 - Token1
 - Hitza1
 - Token2
 - Hitza2
 - ...

7.Bilera-akta

Data: 2015-03-11

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:00

Hitz egin beharrekoa:

- Leihoak

Egin beharrekoa:

- Kotsola bat aurkitu *Elasticsearch* erabiltzeko.
- Bilaketa arruntetan bi eremu desberdinekin saiatu.
- Leihoak

8.Bilera-akta

Data: 2015-03-25

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:15

Hitz egin beharrekoa:

- Leihoak

Egin beharrekoa:

- Eskeman eskatutako aldaketak egin.
- Forora bidaltzeko galdera prestatu.
- Leihoak.

9.Bilera-akta

Data: 2015-04-15

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:10

Hitz egin beharrekoa:

- Bidali nuen galdera forora, baina inork ez du erantzuten.
- Leihoak

Egin beharrekoa:

- Besteen kezkak irakurri.
- Leihoak eskemaren egitura aldatuz egiten saiatu.
 - Hitz bakoitzari posizio zenbakia esleituz.
 - “/forma/lema/kat/azpikat/kasua /forma/lema/kat/azpikat/kasua ...” moduan.

10.Bilera-akta

Data: 2015-05-11

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 12:30

Amaiera ordua: 13:00

Hitz egin beharrekoa:

- Dokumentuen eskemarako beste proposamena entzun.

Proposamena:

- Erabiliko dugun eskema definitu dugu. Hitz bakoitzak aurreko eta hurrengo 5 hitzen analisiak gordeko ditu. Horrez gain, aurreko eta hurrengo esaldiak ere gordeko ditu (Eskema honekin izugarri errepikatzen da informazioa, baina beste edozein modura leihoena lortzen ez dugu lortzen, beraz, erabaki bat hartu beharra zegoen).
- Analisisien formatua: “//forma//lema//kat//azpikat//kasua //forma//lema//kat//azpikat//kasua ...”

Egin beharrekoa:

- Eskema aldatu, pasa dizkidaten fitxategien arabera (fitx1_hitzak).
- Kibanarekin zerbait prestatu.

11.Bilera-akta

Data: 2015-06-03

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 09:30

Amaiera ordua: 10:15

Hitz egin beharrekoa:

- Egindako eskema erakutsi.
- Bilaketak badabiltzala erakutsi (arruntak, leihoak, etab.)
- Kibana erakutsi.
- Galderak:
 - ‘€’ ikurra ez du onartzen.
 - Indize eta motaren izenak aldatu (hitzak eta hitza).

Egin beharrekoa:

- Eskema berria egin eta datu guztiak sartu.
- Bilaketetarako ‘Wildcard’ begiratu.
- Sinboloena begiratu.
- Kibanarekin interfaze bat sortu.

12.Bilera-akta

Data: 2015-06-22

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 09:30

Amaiera ordua: 10:00

Hitz egin beharrekoa:

- Analisiaren barruan hitz konposatuek arazoak ematen dituzte.
- Kibanarekin proposatutako interfazea erakutsi.

Egin beharrekoa:

- `_type` eremua aldatu. 'hitza' balioa jarri denei.
- Kibana interfazea erabiltzailearentzat aldatu.

13.Bilera-akta

Data: 2015-07-15

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 11:30

Amaiera ordua: 12:15

Hitz egin beharrekoa:

- Kibanaren interfazea aldatzea ez da bideragarria.
- Beste eskemaren proposamena egin.

Adostutakoa:

- Prototipoa ez dugu Kibanarekin egingo, ez baita bideragarria bere interfazea aldatzen hastea. Nik nahi dudana teknologia erabili ahal izango dut.
- Prototipo bat egiteko aukerak aztertu. Litekeena da PHP erabiltzea, dagoeneko horren oinarri bat badudalako.

14.Bilera-akta

Data: 2015-09-25

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 11:30

Amaiera ordua: 12:15

Hitz egin beharrekoa:

- PHP erabili dut prototipoa egiteko.
- Bi hitzekin dagoeneko bilaketak egiten ditu. Hori erakutsi ea zer iruditzen zaien.

Adostutakoa:

- PHPrekin aurrera jarraituko dugu.
- Web-aplikazioa osatzeko falta dena egin:
 - Formularioan gehitu:
 - Analisi zuzena edo ez-zuzena nahi duten aukeratzeko atal bat.
 - Hirugarren hitz bat.
 - JSON formatura pasatu formularioko atal guztiak. Formularioak bertatik atzituiko ditu datuak.
 - Hitz bakoitzaren analisia erakusteko atala.

15.Bilera-akta

Data: 2015-11-04

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 09:30

Amaiera ordua: 10:15

Hitz egin beharrekoa:

- Web-aplikazioa bukatuta.

Adostutakoa:

- Orain arte 100 dokumentu besterik ez daude sartuta. Beraz, 10000 fitxategi indexatu denborak kalkulatzeko hasteko.

16.Bilera-akta

Data: 2015-11-25

Kideak: Aitor Soroa, Xabier Artola eta Maialen Otamendi

Lekua: Informatika Fakultatea

Hasiera ordua: 09:30

Amaiera ordua: 10:00

Hitz egin beharrekoa:

- Indexatu beharreko dokumentu guztiak indexatuta daude.

Adostutakoa:

- Memoria osatzeko falta diren atalak idazten hasi.
- Denborak kalkulatu.

Bibliografia

- [1] Paro, Alberto. *Elasticsearch Cookbook*. Packt publishing: Birmingham, 2013.
- [2] *Elasticsearch*en webgune ofiziala: <https://www.elastic.co/>
- [3] Tong, Zachary. *Elasticsearch: The definitive guide*:
<https://www.elastic.co/guide/en/elasticsearch/guide/1.x/index.html>
- [4] Gascón Sánchez, Jagoba (2013). *Armiarma: Balidazio-sistemetan oinarritutako interfaze-sorkuntza dinamikoa*. Gradu Amaierako Proiektua, UPV/EHU: <http://hdl.handle.net/10810/10708>