

## Gradu Amaierako Lana

Informatika Ingeniaritzako Gradua

Konputazioa

---

### **MIR azterketetan erantzun zuzena iragartzeko ezagutza-baseen eta hizkuntza-ereduen erabilpena**

---

*Paula Ontalvilla Gutiérrez*

#### **Zuzendariak**

Oier Lopez de Lacalle

Aitziber Atutxa

2023.eko ekainaren 23



# Esker onak

Hasteko eskerrak eman nahi dizkiet nire tutoreei, Koldo eta Aitziberri, emandako laguntza guztiagatik eta nigan jarritako konfiantza guztiarengatik. Zuengandik asko ikasi dut.

Bilbon eta Donostin ezagutu ditudan klase kideei pasa ditugun momentu guztiengatik. Bereziki, pisu kideei, lau urte hauetan beti egoteagatik. Hasieran klase kide, gero pisu kide eta azkenean, lagun bihurtu zarete.

Azkenik, nire familiari eskerrak eman nahi dizkiet, ama eta ahizpei, emandako babesa, laguntza eta konfiantzagatik. Beti hor egon zarete.

Baina batez ere, nire aititeri, lerro hauek inoiz irakurriko ez badituzu ere.



# Laburpena

Médico Interno Residente (MIR) oso garrantzitsuak diren azterketak dira, Espainian mediku graduatuek egiten dituzte beraiek nahi duten espezialitatean formatzeko. Azterketa horien ebazpenak publikoak ez dira eta gainera, ez dago ia literaturarik MIR azterketetako galderak ebazteko eraikitako sistema desberdinen inguruan. Hori dela eta lan honen helburua MIR azterketak ebazteko sistema bat eraikitzea da.

Horretarako, hasteko, medikuntza ezagutza-base handi, aberats eta eleaniztun bat sortu da ezagutza-base desberdinen informazioa bateratuz. Ondoren, aukera egokia iragartzeko asmoz, metodo desberdinak erabiltzen dituzten bi sistema eraiki dira. Batek, sortutako ezagutza-basean dagoen informazioaz baliatuz erantzun egokia iragarri eta arrazoituko du. Besteak, aldiz, lan bat oinarritzat hartuz, ikasketa automatikoko metodoak, hizkuntza-ereduak batez ere, erabiliko ditu erantzun zuzena iragartzeko. Egindako esperimenduek ataza honen zailtasuna agerian jarri dute. Izan ere, bi hurbilpenekin egindako esperimenduetan behin baino ez da %50eko zehaztasuna gainditu. Hala ere, oinarritzat hartu den ereduaren emaitzak hobetzea lortu da, laneko 0,37ko zehaztasuna eta bere eredu gure datuetan ebaluatuz lortutako 0,44ko zehaztasuna, 0,47ko zehaztasunarekin hobetu da. Bestetik, lanaren mugak gainditu dira. Eraikitako eredu berri baten bidez 5 erantzun dituzten galderak ebaztea lortu da eta ezagutza-baseekin horretaz aparte, beste hizkuntza batzuetan idatzita dauden azterketak ebaztea lortu da. Gainera, ezagutza-base oso baliagarri eta aberatsa sortu da aplikazio desberdinak izan ditzakeenak etorkizunean.



# Gaien aurkibidea

<b>Laburpena</b>	<b>i</b>
<b>Gaien aurkibidea</b>	<b>iii</b>
<b>Irudien aurkibidea</b>	<b>v</b>
<b>Taulen aurkibidea</b>	<b>vii</b>
<b>1 Sarrera</b>	<b>1</b>
1.1 Motibazioa . . . . .	1
1.2 Proiektuaren helburuak . . . . .	2
<b>2 Proiektuaren plangintza</b>	<b>5</b>
2.1 Lanaren deskonposaketa egitura (LDE) . . . . .	5
2.1.1 Lan-paketeak . . . . .	5
2.2 Denbora kudeaketa . . . . .	7
2.2.1 Denbora estimazioa . . . . .	7
2.2.2 Gantt diagrama . . . . .	7
2.3 Emangarriak eta mugarriak . . . . .	8
2.4 Komunikazio-plana . . . . .	8
2.4.1 Komunikazio-kanalak . . . . .	10
2.4.2 Informazio sistema . . . . .	10
2.5 Arriskuen analisisa . . . . .	10
2.6 Jarraipen eta kontrola . . . . .	11
2.6.1 Aurkitutako arazoak eta plangintzaren egokitzapenak . . . . .	11
2.6.2 Plangintzarekiko desbiderapenak . . . . .	12
<b>3 Oinarri teorikoak eta aurrekariak</b>	<b>13</b>
3.1 Oinarri teorikoak . . . . .	14
3.1.1 Hurbilpenak . . . . .	14
3.1.2 Hizkuntzaren Prozesamenduko atazak . . . . .	19
3.2 Aurrekariak . . . . .	21
3.2.1 NER medikuntza domeinuan . . . . .	22
3.2.2 Medical multiple-choice question answering (MedMCQA) . . . . .	23
<b>4 Garatutako implementazioa</b>	<b>25</b>
4.1 Datuak . . . . .	25
4.1.1 CasiMédicos-MIR . . . . .	25

4.1.2	MedMCQA <i>dataset</i> -a . . . . .	26
4.1.3	Sortutako ezagutza-basea . . . . .	27
4.2	Metrikak . . . . .	33
4.2.1	Ezagutza-baseetan oinarritutako hurbilpeneko metrikak . . . . .	34
4.2.2	MCQA hurbilpeneko metrikak . . . . .	34
4.3	Ezagutza-baseetan oinarritutako hurbilpena . . . . .	34
4.3.1	Funtzionalitateak eta hauen inplementazioa . . . . .	35
4.4	MCQA hurbilpena . . . . .	42
4.4.1	Sistemaren konfigurazioa . . . . .	42
4.4.2	Oinarrizko eredia . . . . .	42
4.4.3	Egindako esperimentuak . . . . .	44
<b>5</b>	<b>Emaitzak eta eztabaida</b> . . . . .	<b>51</b>
5.1	Ezagutza-baseetan oinarritutako hurbilpena . . . . .	51
5.2	MCQA hurbilpena . . . . .	55
5.2.1	Oinarrizko eredia (bert-base-uncased + MedMCQA) . . . . .	55
5.2.2	Oinarrizko ereduaren <i>bert</i> -a CasiMédicos-MIR corpusarekin findu (bert-base-uncased + CasiMédicos-MIR ) . . . . .	55
5.2.3	Oinarrizko eredia MedMCQA dataset-arekin eta CasiMédicos-MIR (bert-base-uncased + MedMCQA + CasiMédicos-MIR) . . . . .	58
5.2.4	Bost erantzuneko eredu berria . . . . .	60
5.2.5	Ereduen arteko konparazioa . . . . .	61
5.3	Bi hurbilpenen konparazioa . . . . .	61
<b>6</b>	<b>Ondorioak eta etorkizunerako lana</b> . . . . .	<b>63</b>
6.1	Etorkizunerako lana . . . . .	64
<b>A</b>	<b>MCQA hurbilpeneko emaitzak</b> . . . . .	<b>1</b>
A.1	Oinarrizko ereduaren <i>bert</i> -a CasiMédicos-MIR corpusarekin findu (bert-base-uncased + CasiMédicos-MIR ) . . . . .	1
A.1.1	4 aukera posible dituzten galderen esperimentuak . . . . .	1
A.1.2	4 eta 5 aukera posible dituzten galderen esperimentuak . . . . .	1
A.2	Hurbilpenen konparazioa . . . . .	1
	<b>Bibliografia</b> . . . . .	<b>5</b>



# Irudien aurkibidea

1.1	Medikuntzarako ezagutza-base eleanitza eta aberatsa lortzeko sistema. . . . .	3
1.2	MIR azterketen erantzun egokia iragartzeko sistemak. . . . .	3
2.1	Lanaren deskonposaketa egitura diagrama. . . . .	5
2.2	Lanaren Gantt diagrama. . . . .	9
3.1	Biomedikuntza domeinuan Wikidatan dagoen kontzeptu kopurua, motaren arabera. Turki et al.-ren artikulutik aterata . . . . .	17
3.2	Transformerraren arkitektura [2]. artikulutik hartua. . . . .	18
3.3	[3]. artikulutik aterata, BERT-aren aurre-entrenamendua eta <i>fine-tuning</i> -a. . .	19
3.4	NER etiketatzearen adibide bat. Goian <i>BIO</i> etiketatzerik gabe, behean <i>BIO</i> etiketatzearekin. . . . .	20
3.5	DREAM <i>dataset</i> -eko adibide bat Jin et al.-ren artikulutik aterata dagoena. ✓: erantzun egokia adierazten du. . . . .	21
3.6	Goiko esaldian laranja Metamap-ek, berdez cTAKES-ek eta morez biek detektatuko lituzketen entitate medikoak, Nixon et al.-ren lanatik ateratako adibidea. Beheko esaldian UNIMER-ek bueltatutako laranja detektatutako sintomak eta morez detektatutako gaixotasunak . . . . .	23
4.1	Ezagutza-basea sortzeko egin den sistema. . . . .	27
4.2	Wikipediako ingelesezko <i>arthritis</i> gaixotasunaren <i>infobox</i> -a. . . . .	28
4.3	Hasieran ezagutza-baseak zituen gaixotasunak eta Wikipediarekin eta BabelNet-ekin estaldura handitu ostean lortutako gaixotasun kopurua hizkuntza bakoitzeko. . . . .	33
4.4	Adibide bat, galderan urdinez dauden sintomak eta gorritz, erantzunetan detektatutako gaixotasunak daude. Erantzunen azpian gaixotasun horiek eraikitako ezagutza-basean bilatuz, dituzten sintomak daude. Beltzez komunean dituzten sintomak daude, lehenengoak gehiago dituzenez, hura zuzena da. . . . .	35
4.5	Gaixotasunak eta sintomak identifikatzeko sistema. . . . .	35
4.6	Gaixotasunak eraikitako ezagutza-basean bilatzeko sistema. . . . .	38
4.7	Aukera egokia iragartzeko sistema. . . . .	41
4.8	Oinarrizko ereduaren arkitektura [6]. . . . .	44
4.9	Eredu berriaren arkitektura, 5 erantzun posible dituzten galderak erantzuteko gai dena. . . . .	49
5.1	Asmatutako bi galdera eta bere erantzun posibleak, urdinez galderetan aurkitu diren sintomak eta gorritz erantzunetan aurkitu diren gaixotasunak agertzen dira. Erantzun bakoitzeko gaixotasunen sintomak ere agertzen dira komunean dituzten sintomak beltzez daudelarik. ✓: erantzun egokia adierazten du. . . .	53

5.2	Ingeleseztan ingelesez NCI bidez detektatu diren sintomak. Urdinez, galderan detektatutako sintomak eta berdez, NCI bidez lortutako sintoma berriak. Beltzez, komunean dituzten sintomak. NCI bidez, <i>fever</i> birritan agertzen denez, bi erantzunen artean berdinketa dago. . . . .	54
-----	--	----

# Taulen aurkibidea

2.1	Lan-pakete bakoitzeko aurreikusitako denbora. . . . .	8
2.2	Proiektuaren mugarrak. . . . .	10
2.3	Lan-pakete bakoitzeko aurreikusitako denbora eta errealitatean erabili den denbora. . . . .	12
3.1	Ezagutza-base desberdinekin gordetzen den informazioa <i>Arthritis</i> eta <i>Alcoholic liver disease</i> gaixotasun bakoitzarekin. . . . .	15
4.1	CasiMédicos-MIR <i>dataset</i> -ean gaztelaniaz dauden bi kasuren adibidea. Lehenengoak 5 aukera ditu eta bigarrenak, aldiz, 4 bakarrik. Hutsik zeuden zutabeak ezabatu dira. . . . .	26
4.2	MedMCQA <i>dataset</i> -aren adibide bat. . . . .	26
4.3	Wikidata, DBPedia eta Wikipedia arteko konparazioa. <i>Arthritis</i> eta <i>Alcoholic liver disease</i> gaixotasun bakoitzarekin lor daitezkeen propietateak. Horiek parakatuta daude iturri desberdinen beste propietateekin eta beltzez interesatzen zaizkigun propietateak daude. . . . .	29
4.4	BabelNet-etik lortutako gaixotasunen inguruko emaitzak. . . . .	32
4.5	MetaMap-etik eta UNIMER-retik ingelesez eta gaztelaniazko galderak eta erantzunak pasa ondoren lortutako gaixotasun eta sintoma <b>kode</b> desberdinak. . . . .	37
4.6	MetaMap-etik eta UNIMER-retik ingelesez eta gaztelaniazko galderak eta erantzunak pasa ondoren komunean eta komunean ez dituzten gaixotasun eta sintoma kodeak. . . . .	38
4.7	Ingelesez eta gaztelaniaz zenbat kasu daude zeinetan entitateak detektatu dituzten erantzun kopurua 0, 1, 2, 3, 4 eta 5 den, hurrenez hurren. . . . .	38
4.8	Eraikitako ezagutza-basean aurkitu diren erantzunetako gaixotasun kodeak eta lortutako sintomak ingelesez eta gaztelaniaz. . . . .	39
4.9	Eraikitako ezagutza-basean Levenshtein distantzia bidez aurkitutako erantzunetako gaixotasunen izenak eta lortutako sintomak ingelesez eta gaztelaniaz. <u>Azpimarratuta</u> UNIMER-rek gaixotasun edo sintoma moduan identifikatu eta kodea lortu ezin izan den gaixotasun entitateen izenekin lortutako emaitzak. Gaztelaniaz bigarren errenkadan <b>erlazioekin</b> lortutakoa emaitzak daude. . . . .	40
4.10	Eraikitako ezagutza-basean aurkitu diren erantzunetako gaixotasun kodeak eta lortutako sintomak ingelesez eta ingelesez NCI bidez lortutako sintomekin eraikitako <i>csv</i> -tan bilatuta lortutakoak. . . . .	40
4.11	Eraikitako ezagutza-basean Levenshtein distantzia bidez aurkitutako erantzunetako gaixotasunen izenak eta lortutako sintomak ingelesez eta ingelesez NCI bidez lortutako sintomekin eraikitako <i>csv</i> -tan bilatuta lortutakoak. . . . .	40
4.12	MedMCQA <i>dataset</i> -aren adibide bat. . . . .	44

4.13	CasiMédicos-MIR corpuseko <i>test</i> partizioan dauden kasu kopurua egoera desberdinetan. . . . .	46
4.14	CasiMédicos-MIR <i>train</i> eta <i>dev</i> partizioetan dauden kasu kopurua egoera desberdinetan. . . . .	47
4.15	MedMCQA-k duen egitura berdina izan dezan, aldatutako CasiMédicos-MIR-eko corpuseko adibide bat. . . . .	50
5.1	Ezagutza-baseetan oinarritutako hurbilpenean galdera <b>guztien</b> artean lortutako zehaztasuna *100 eginez hizkuntza desberdinetan. Zehaztasuna era desberdinetan kalkulatu da eta parentesi artean asmatutako erantzun kopurua agertzen da. . . . .	52
5.2	Hizkuntza bakoitzeko bueltatutako emaitzen galderen inguruko informazioa. .	54
5.3	Sistemak bueltatutako erantzunak bakarrik kontuan harturik lortutako zehaztasuna * 100 eginez hizkuntza desberdinetan. Parentesi artean asmatutako galdera kopurua agertzen da. . . . .	54
5.4	Oinarrizko ereduarekin (bert-base-uncased + MedMCQA) lortutako zehaztasuna.	55
5.5	Oinarrizko ereduak erabiltzen duen <i>bert-base-uncased</i> CasiMédicos-MIR 4 aukera dituzten galderekin finduz lortutako emaitzen batezbestekoa eta desbiderapen estandarra. . . . .	56
5.6	Oinarrizko ereduak erabiltzen duen <i>bert-base-uncased</i> CasiMédicos-MIR corpus osoarekin, 4 eta 5 aukera dituzten galderekin, finduz lortutako batezbestekoa eta desbiderapen estandarra. . . . .	58
5.7	Oinarrizko ereduak MedMCQA dataset-arekin eta CasiMédicos-MIR corpusarekin finduz lortutako zehaztasuna. Beltzez emaitzak hobetzea lortu diren esperimentuak eta emaitzak. . . . .	59
5.8	Sortutako eredu berriarekin lortutako zehaztasun emaitzak. . . . .	61
5.9	Ezagutza-base hurbilpenak emaitza bat bueltatu dituen galderak <b>bakarrik</b> kontuan harturik lortutako zehaztasuna esperimentu desberdinekin. . . . .	62
A.1	Oinarrizko ereduaren <i>bert</i> -a CasiMédicos-MIR corpusarekin finduz (bert-base-uncased + CasiMédicos-MIR) lortutako emaitzak <i>seed</i> desberdinekin eta batezbestekoa eta desbiderapen estandarraren kalkulua. . . . .	2
A.2	Oinarrizko ereduaren <i>bert</i> -a CasiMédicos-MIR corpus osoarekin, 4 eta 5 aukera dituzten galderekin, finduz lortutako emaitzak <i>seed</i> desberdinekin eta batezbestekoa eta desbiderapen estandarraren kalkulua. . . . .	3
A.3	Ezagutza-base hurbilpenak erantzun bat bueltatutako galderak bakarrik kontuan harturik lortutako zehaztasuna esperimentu desberdinekin. . . . .	3

# Sarrera

## 1.1 Motibazioa

Medikuntza gero eta informatizatuago dagoen arlo bat da. Teknologia berriek medikuntza erregistro elektronikoak sortzea, denbora errealean datu klinikoak sartu eta eskuratzea, administrazio atazak automatizatzea, etab. ekarri dute. Honi esker, prozesu klinikoak hobetzea eta azkartzea lortu da eta osasun arloko adituen arteko koordinazio hobetu da. Izan ere, pazientearen informazioa eskuratzea azkarragoa da, diagnostiko eta tratamendu prozesuak optimoagoak eta hartutako erabakiak hobeak eta zehatzagoak dira. Gainera, telemedikuntzak aukera eman du medikuntza zerbitzuak urrunetik eskaintzeko, urruneko eremuetan edo muga geografikoak dituztenetan arreta jasotzeko aukera erraztuz.

Bestetik, Adimen Artifiziala (AA) era askotan ezarri da gure eguneroko bizitzan, hala nola laguntzaile birtualak (Siri, Alexa...), garraio, hegazkingintza eta joko informatiko automatikoak, etab. Berrikiago, informatizazioari esker, AA medikuntzan sartzen hasi da. Horrela, pazientearen arreta hobetzeko prozesuak azkartu dira eta zehaztasuna handiagotu da. Irudi erradiologikoak, patologia-diapositibak eta pazienteen historia kliniko elektronikoak (*Electronical Medical Resource*, EMR) ikasketa automatikoa erabiliz ebaluatzen hasi dira, pazienteen diagnostikoaren eta tratamenduaren prozesuan lagunduz eta medikuen gaitasunak areagotuz. Oro har, AA osasun hobea eskaintzeko bidea irekitzen ari da [7].

AA-k datu-multzoen beharra izaten du ataza desberdinak egin ahal izateko, adibidez, datuetan patroiak identifikatu ahal izateko. Datu horiek alde batetik, informatizazioari esker areagotu dira, pazienteen inguruko informazioa, txosten klinikoak, erregistro elektronikoak, etab. gordetzea ahalbidetu duelako. Beste alde batetik, informatizazioak ikerketa sustatu duenez, medikuntzako literatura azken urteotan asko hazi da. Gainera gero eta ontologia, hiztegi eta ezagutza-base gehiago sortu dira medikuntza termino desberdinak gordetzen dituztenak. Horiek funtsezkoak dira kontzeptu medikoak era egituratu batean sailkatu eta antolatzeke, sistema informatikoei erabilera erraztuz.

Eskuragarri dauden datu medikoen artean medikuntzako azterketak ere aurkitzen dira. Espainia mailan medikuek egin behar duten azterketa ezagunena MIR (Médico Interno Residente) da. Azterketa hori, test motatakoa dena, mediku graduatuek egiten dute beraiek nahi duten espezialitatean formatzeko. Nota altuena lortu duenetik txikiena lortu duena

arte, ordenean, aukeratzen da espezialitatea. Zenbat eta nota altuago, orduan eta hobeto, bakoitzak nahi duen espezialitatea aukeratzeko aukerak handiagoak direlako. Hori dela eta, oso azterketa garrantzitsua da.

MIR azterketak eta horien ebazpenak ez dira publikoak. Horiek prestatzeko sortu diren akademiek MIR Asturias, CTO, AMIR, PROMIR esate baterako, izaten ohi dituzte. Hala ere, CasiMédicos webguneak MIR azterketa ebatziak doan igota ditu mediku eta graduatu bolondresez ebatziak eta azalduta. Gainera, ez dago ia literaturarik MIR azterketetako galderak ebazteko eraikitako sistema desberdinen inguruan. Izan ere, [Vilares and Gómez-Rodríguez](#)-ren lan baino ez da aurkitu; bertan, besteak beste MIR azterketak dituen datu-multzo bat eraiki eta ebatzi egiten da. Hala ere, beste medikuntzako test azterketak dituzten datu-multzo desberdinak sortu eta ebazteko sistemak aurkitu daitezke literaturan.

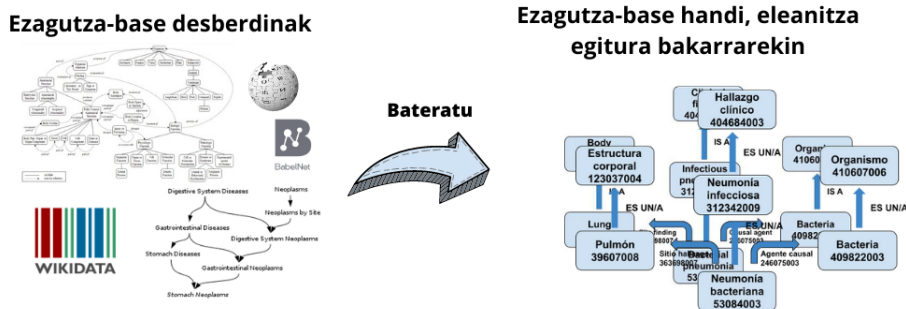
Aipatutako guztiagatik, MIR azterketak ebazteko sistema bat eraikitzearen beharra ikusi da zeinak erantzun egokia iragarri eta justifikatzeko gai den. Lan honetan, hasteko, medikuntzarako ezagutza-base handi, aberats eta eleaniztuna sortu da. Ondoren, aukera egokia iragartzeko bi sistema eraiki dira bakoitzak metodo desberdinak erabiltzen dituelarik. Batek, sortu den ezagutza-basean dagoen informazioaz baliatuz erantzun egokia iragarri eta arrazoituko du eta besteak, aldiz, ikasketa automatikoki metodoak erabiliko ditu erantzun zuzena iragartzeko.

Egindako lana ondo ulertzeko, [3.](#) kapituluan beharrezkoak diren oinarri teorikoak eta arloaren egoera aztertuko da, arloan erabilitako teknikak eta lortutako emaitzak aurkeztuz. Horrela, [4.](#) kapituluan azaldutako inplementazioa eta [5.](#) kapituluan lortutako emaitzak ulertu ahal izango dira. Bestalde, [2.](#) kapituluan proiektua aurrera eramateko jarraitutako plangintza aurkezten da. Azkenik, [6.](#) kapituluan ateratako ondorioak laburtzen dira, eta ondorio horietan oinarrituta etorkizunerako lana proposatzen da. Arlo zientifikoari dago-kionez, proiektu honetan testu idatziekin lana egingo denez, hizkuntzaren prozesamenduko arlo desberdinak lantzen dira.

## 1.2 Proiektuaren helburuak

Proiektu honen helburu nagusia Médico Interno Residente (MIR) azterketetako galderak eta erantzun posibleak izanda, galderak modu egokian erantzuteko gai den sistema bat eraikitzea da. Helburu nagusi horretara ailegatzeko beste hainbat helburu definitu dira:

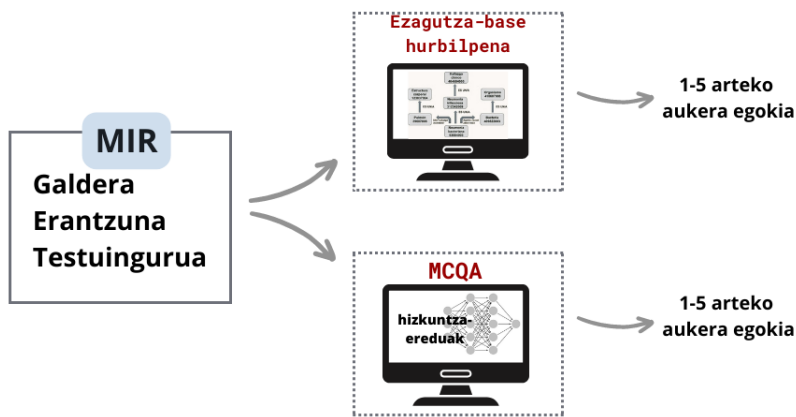
- Medikuntzarako ezagutza-base eleanitza eta aberatsa lortzea. Ezagutza-base desberdinek medikuntzako informazio asko bilduta dute bakoitzak egitura desberdin bat daukalarik. Gainera askok beraien artean osagarri den informazioa daukate. Hori dela eta, informazio iturri desberdinetatik medikuntzako informazioa erauzi eta ezagutza-base aberats bakar bat sortuko duen sistema bat egingo da egitura zehatz bat izango duena. Ezagutza-base honek hizkuntza desberdinetan izango du informazioa gordeta. Egingo denaren adierazpen grafikoa [1.1.](#) irudian ikus daiteke.
- MIR azterketako galderen erantzun egokia iragarriko duen sistema sortzea. Hori bi modu desberdinetan egin daiteke. Jarraian azalduko dira, baina [1.2.](#) irudian egingo denaren laburpena ikus daiteke:
  1. Ezagutza-baseetan gordetako informazioa erabiliz. Horretarako, sortu den baliabide eleanitz eta aberatsa erabiliz, erantzun egokia iragarriko duen sistema



**1.1 Irudia:** Medikuntzarako ezagutza-base eleanitza eta aberatsa lortzeko sistema.

bat eraiki behar da. Sistema honek ez du aukera zuzena zein den bakarrik iragarriko, erantzun zuzena zergatik egokia den ere azalduko du.

2. Hizkuntza-ereduez baliatuz. Horretarako, Hizkuntzaren Prozesamenduko *Multiple Choice Question Answering (MCQA)* ataza medikuntza arloan aplikatuko duen sistema bat eraiki behar da. Honek hizkuntza-ereduak eta medikuntzako corpus bat erabiliko ditu. Eraikiko den sistemak Pal et al.-ren lana erreferentzia moduan hartuko du eta aldaketak egingo dizkio. Horrela, hizkuntza-ereduen eta medikuntzako inguruko jakintzaz baliatuko da erantzun zuzena zein den iragartzeko. Baina ezingo da jakin sistemak aukeratu duen erantzunaren justifikazioa.



**1.2 Irudia:** MIR azterketen erantzun egokia iragartzeko sistemak.



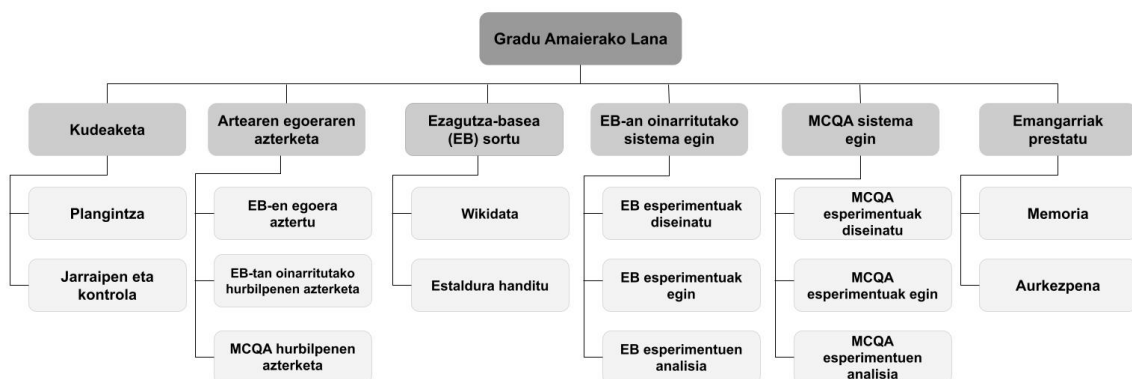


# Proiektuaren plangintza

Kapitulu honetan, lanaren deskonposaketa egitura (LDE) eta definitutako lan-paketeen zerrenda emango dira. Honetaz gain, mugarrak, aurreikusitako denbora eta Gantt diagrama aurkeztuko dira. Gainera, arrisku-plana eta proiektuan egin den jarraipen eta kontrola azalduko dira.

## 2.1 Lanaren deskonposaketa egitura (LDE)

Proiektuaren helburuak bete daitezten lanaren deskonposaketa egitura (LDE) sortu da, 2.1. irudian aurkezten dena. Bertan sortutako adarrak eta bakoitzean definitu diren lan-paketeak ikus daitezke. Jarraian, definitutako adarrak eta horiek osatzen dituzten lan-paketeen deskribapena egiten da.



2.1 Irudia: Lanaren deskonposaketa egitura diagrama.

### 2.1.1 Lan-paketeak

#### Kudeaketa adarra

Adar honetan proiektuaren kudeaketarekin zerikusia duten paketeak egongo dira.

- **Plangintza garatu (K1).** Pakete honek hasierako plangintza egiteko atazak, hala nola lan-paketeak zehaztea, denbora estimazioak, Gantt diagrama egitea, arrisku-plana eta lan-metodologia zehaztea barnean izango ditu. Horrez gain, behar badira, plangintza egunean mantentzeko behar daitezkeen atazak ere barne izango ditu.
- **Jarraipena eta kontrola (K2).** Lan-pakete honek proiektuaren garapen egokia bermatuko duten atazak edukiko ditu, hau da, dedikazioen jarraipena eta epeen eta espezifikazioen betetzea, arriskuen kudeaketa eta egindako bilerak.

### Artearen egoeraren azterketa adarra

Adar honetan artearen egoera aztertuko duten lan-paketeak daude. Arteari buruzko egoera aztertzea ezinbestekoa izango da egindako lanak artean ekarpen bat egingo duela konprobatzeko. Era berean, implementazioan eragina izango du.

- **Ezagutza-baseen (EB) egoera aztertu (AE1).** Medikuntzako ezagutza-base desberdinak eta horiek sortzeko erabilitako teknika, iturri eta baliabide desberdinak aztertu eta bilatzeko atazak sartuko dira.
- **EB-tan oinarritutako hurbilpenen azterketa (AE2).** Ezagutza-baseetan oinarrituta eta batez ere, Wikidatan oinarrituta egin diren antzeko hurbilpenak eta metodo desberdinak aztertzen dituzten atazak sartuko dira.
- **MCQA hurbilpenen azterketa (AE3).** Medikuntza arloan MCQA-n dauden *dataset*, teknika eta metodo desberdinak aztertuko dituzten atazak barnean izango ditu.

### Ezagutza-basea (EB) sortu adarra

Hemen ezagutza-basea sortuko duen sistema diseinatu eta garatzeko lan-paketeak egongo dira.

- **Wikidata (EB1).** Wikidatatik informazioa erauzi eta egituratzeko atazak sartzen dira.
- **Estaldura handitu (EB2).** Ezagutza-basearen estaldura handitzeko egingo diren atazak ditu barnean.

### EB-etan oinarritutako sistema egin adarra

Adar honetan ezagutza-baseetan oinarritutako hurbilpena egiteko zerikusi duten lan-paketeak daude.

- **EB esperimentuak diseinatu (EBH1).** Ezagutza-basea sortuta eta artearen egoera aztertuta dagoela, egingo diren esperimentuak pentsatzeko eta prestatzeko atazak sartuko dira.
- **EB esperimentuak egin (EBH2).** Esperimentuak prestatu ondoren, horiek aurrera eramateko egindako atazak barnean izango ditu.
- **EB esperimentuen analisia (EBH3).** Esperimentuak amaitu ostean lortutako emaitzak aztertzeko eta interpretatzeko atazak sartuko dira.

### MCQA sistema egin adarra

Adar honetan MCQA hurbilpena egiteko zerikusi duten lan-paketeak daude.

- **MCQA esperimentuak diseinatu (MCQA1).** Artearen egoera zein den eta bertan erabilitako teknikak eta *dataset*-ak aztertu direla, egingo diren esperimentuak pentsatzeko eta prestatzeko atazak sartzen dira.
- **MCQA esperimentuak egin (MCQA2).** MCQA esperimentuak prestatu ondoren, horiek aurrera eramateko egindako atazak barnean izango ditu.
- **MCQA esperimentuen analisia (MCQA3).** MCQA esperimentuak amaitu ostean lortutako emaitzak aztertzeko eta interpretatzeko atazak sartuko dira.

### Emangarriak prestatu adarra

Adar honetan Gradu Amaierako Lanean entregatu behar diren bi emangarriak prestatzeko lan-paketeak daude.

- **Memoria (E1).** Lan-pakete honek memoria idazteko, zuzentzeko eta entregatzeko atazak barnean hartuko ditu.
- **Aurkezpena (E2).** Hemen aurkezpena prestatzeko atazak, adibidez, gidoia egin edota gardenkiak prestatu sartuko dira.

## 2.2 Denbora kudeaketa

Proiektua bi urteetan zehar garatua izan da. 2021/2022 kurtsoan, 2022/07/11 EHAA-ko deialdiko erresoluzioan argitaratua izan zen Ikasiker bekarekin ezagutza-basea sortu zen. 2022/2023 kurtsoan, aldiz, Hizkuntzaren Prozesamenduaren inguruko ezagutza nahiko izanda, gainerako lana egin da.

Nahiz eta Gradu Amaierako Lanak ikasketa-planeko 12 ECTS balio dituen eta ondorioz, 300 lan ordukoa izatea aurreikusi, lan honetan aurreikusi egin den denbora gehiago izan da. Hori kontuan izanik, atal honetan proiektua aurrera eramateko helburuarekin, lan-pakete bakoitza egiteko aurreikusi den denbora eta garapen denboraldiak aurkeztuko dira.

### 2.2.1 Denbora estimazioa

Lan-pakete bakoitza egiteko estimatutako denbora [2.1.](#) taulan aurkitzen da.

### 2.2.2 Gantt diagrama

Lan-pakete bakoitza denboran noiz garatuko den [2.2.](#) irudian agertzen den Gantt diagraman aurkezten da. Lehenik eta behin, ezagutza-basea sortzeko plangintza egingo da eta ondoren ezagutza-basea sortuko da. Horretarako, horren artearen egoera aztertu, eta gero Wikidata erabiliz ezagutza-basea sortuko da (EB1) eta jarraian, estaldura handituko da (EB2).

Aipatutakoa amaitu ondoren, urte bateko itxaronaldia egongo da. Denbora hori pasa ostean, egiteke geratzen diren lan-paketeen plangintza egingo da. Jarraitzeko, “EB-an oinarritutako sistema egin” adarra garatuko da. Hori dela eta, artearen egoera aztertu

Lan-paketeak	Ordu kopurua
<b>Kudeaketa</b>	<b>80</b>
Plangintza garatu	30
Jarraipena eta kontrola	50
<b>Artearen egoeraren azterketa</b>	<b>80</b>
EB-en egoera aztertu	30
EB-tan oinarritutako hurbilpenen azterketa	20
MCQA hurbilpenen azterketa	30
<b>Ezagutza-basea (EB) sortu</b>	<b>100</b>
Wikidata	70
Estaldura handitu	30
<b>EB-an oinarritutako sistema egin</b>	<b>60</b>
EB esperimentuak diseinatu	5
EB esperimentuak egin	30
EB esperimentuen analisia	25
<b>MCQA sistema egin</b>	<b>80</b>
MCQA esperimentuak diseinatu	15
MCQA esperimentuak egin	50
MCQA esperimentuen analisia	15
<b>Emangarriak prestatu</b>	<b>75</b>
Memoria	60
Aurkezpena	15
<b>Guztira</b>	<b>475</b>

2.1 Taula: Lan-pakete bakoitzeko aurreikusitako denbora.

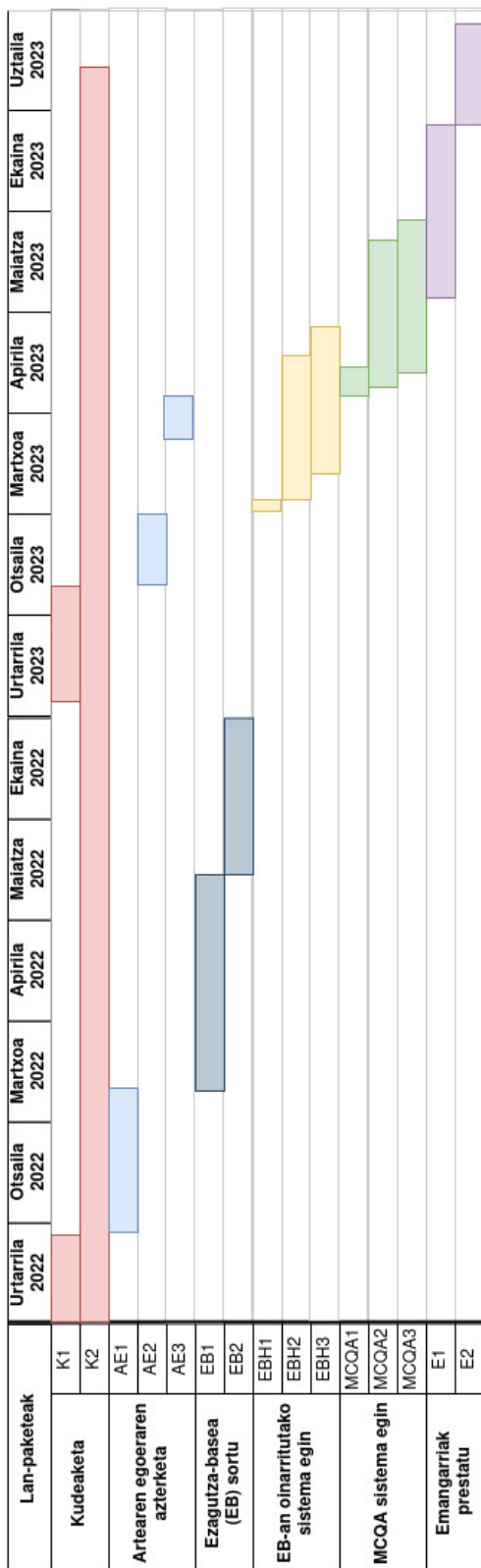
(AE2), esperimentuak diseinatu (EBH1) eta ia paraleloan, esperimentuak egin (EBH2) eta esperimentuen emaitzak analizatuko dira (EBH3), hau da, esperimentuak amaitu ahala horien emaitzak aztertuko dira. Bigarrenik, nahiz eta aurreko adarra ez amaitu, paraleloan, “MCQA sistema egin” adarra garatzen hasiko da. Hortaz, artearen egoera aztertu (AE3), esperimentuak diseinatu (MCQA1) eta berriz ere, ia paraleloan, esperimentuak egin (MCQA2) eta esperimentuen emaitzak analizatuko dira (MCQA3). Azkenik, memoria eta aurkezpena prestatuko dira. Momentu guztietan jarraipen eta kontrola egingo da.

### 2.3 Emangarriak eta mugarriak

Proiektu honek bi emangarri nagusi ditu, memoria eta defentsa aurkezpena, bakoitzak epe-muga desberdin bat daukalarik. Hala ere, biak entregatu baino lehen GrAL-aren matrikula eta defentsaren eskaera egin behar da. Gainera, GrAL entregatu arte ez da jakingo noiz egingo den defentsa beraz, ezin izango da data zehatz bat finkatu plangintzan. Gauzak horrela, 2.2. taulan agertzen dira proiektuaren mugarriak.

### 2.4 Komunikazio-plana

Atal honetan komunikazio-plana azalduko da, zehazki, proiektuko informazio sistema eta komunikazio sistemaren berezitasunak azalduko dira.



2.2 Irudia: Lanaren Gantt diagrama.

Entregak	Epe-muga
GrAL-aren matrikula eta defentsaren eskaera	2023/06/19
Memoriaren entrega	2023/06/25
Proiektuaren defentsa	2023/06/3-14

2.2 Taula: Proiektuaren mugarriak.

### 2.4.1 Komunikazio-kanalak

Proiektuan dauden interesatuen arteko komunikazio bi kanal bidez egingo da:

- **Bilerak.** Tutoreekin bilerak astero egingo dira *webex* bidez. Biler horietan proiektuaren jarraipena egingo da. Era berean, biler horietan proiektuaren eta esperimentuen norabideari buruzko erabakiak hartuko dira.
- **Posta elektronikoa.** Posta elektronikoa erabiliko da kideen artean komunikatzeko, bileren datak zehazteko edo aldatzeko eta garrantzia txikiko kontsultak argitzeko.

### 2.4.2 Informazio sistema

Proiektuaren eduki desberdinak digitalki gordeko dira hurrengo gordailu desberdinetan:

- **IXA taldearen zerbitzariak.** Hizkuntza-eredu eta corpus handiekin lana egingo denez IXA-ko zerbitzariak dituzten GPU-etan exekutatzea erabaki da. Bertan gordeko dira inplementazioaren *script* eta irteera gehienak. Gainera hor segurutasun kopiak egiten direnez, galtzeko arriskua txikia da.
- **Etxeko ordenagailua.** Kostu konputazional txikia duten algoritmoak exekutatzeko ez dira IXA-ko zerbitzariak erabiliko, baizik eta ordenagailu pertsonalean egingo da. Hor segurutasun kopiak egiten ez direnez, datuak galtzeko arriskua altuagoa da.
- **Google Drive.** Bileren aktak eta garrantzi txikiko dokumentuak sortu eta konpartitzeko erabiliko da.
- **Overleaf.**  $\LaTeX$ -en memoria idazteko erabiliko den zerbitzua da, kideen artean konpartitu eta aldi berean editatzen uzten du.
- **GitHub.** Proiektuaren kode guztia gordetzeko eta proiektuaren bertsioak kudeatzeko erabiliko da.
- **WandB.** Entrenatutako modeloen jarraipena egiteko erabiliko da. Honek modeloen hiperparametroak, kostu eta galera funtzioen kurbak, GPU-en erabilera, etab. gordetzeko aukera ematen du.

## 2.5 Arriskuen analisisa

Plangintzan arriskuen identifikazio eta analisisa ezinbesteko atala da. Horregatik proiektu honetan topa ditzakegun arriskuak identifikatu, analizatu eta soluzio posibleak bilatu ditugu. Proiektua irismen handikoa izanik, hainbat alor desberdinetako arriskuak aurreikusi dira.

- Gaixotasun nahiko ez dituen ezagutza-basea eraikitzea. Nahiz eta ezagutza asko egon, hura lortzeko prozesua erraza denaren ziurtasuna ez dago eta gerta daiteke gaixotasun asko lortu ezin izatea. Estaldura handitzen saiatu ostean hori gertatzen baldin bada, ezagutza-baseetan oinarritutako hurbilpenean erabiliko den ezagutza-basea aldatzea aztertuko da. Bestela, horrekin jarraituko da eta lortutako emaitzen ondorioetan islatuko da estaldura txikiaren eragina.
- Arazo teknikoak direla eta Google Drive, Overleaf, IXA zerbitzarietan edo etxeko ordenagailuan dagoen informazioa galdu daiteke. Hori gertatzearen probabilitatea nahiko baxua da horiek segurtasun-kopiak egiten dituztelako. Gainera, etxeko ordenagailuaren informazioa GitHub-en gordeko da segurtasun kopia moduan. Hala ere, datuen galera bat egoten bada gordetako azken puntutik hasiko da lana. Galdutakoa berregitea ezinezkoa bada plangintza egokituko da.
- Proiektu handia izanik plangintzarekiko desbiderapenak gerta daitezke. Arrisku honen ondorioz beste ataza batzuetan atzerapenak egon daitezke. Hori dela eta, plangintzan atazei benetan pentsatzen den denbora baino gehiago esleitu zaie.
- Lortutako emaitzak esperotakoak ez izatea. Ezagutza-baseetan oinarritutako hurbilpen berdina aurretik egin ez denez, gerta daiteke espero diren emaitzak ez lortzea, adibidez Wikidatan gaixotasunak ez aurkitzeagatik. MCQA hurbilpenean ere gerta daiteke lortutako emaitzak esperotakoak ez izatea, baina aurretik egindako lan bat oinarritzat hartu denez hori gertatzearen probabilitate txikiagoa da. Hala ere, esandakoa gertatuz gero, aztertuko litzateke zer nolako eragina daukan proiektuan eta nola konpondu daitekeen plangintza egokitzuz. Betiere, memorian jasoko dira lortutako emaitzak eta ondorioak.

## 2.6 Jarraipen eta kontrola

Atal honetan lanaren jarraipenean zehar aurkitutako arazoak eta horiei aurre egiteko hartutako neurriak azalduko dira. Gainera, lan-pakete bakoitza egiteko aurreikusitako denboraren eta benetan erabili den denboraren arteko desbiderapenak aztertuko dira.

### 2.6.1 Aurkitutako arazoak eta plangintzaren egokitzapenak

Lanean izandako arazo nagusia esperotako emaitzak ez lortzea izan da, arriskueta aurreikusi zen bezala. Eraikitako ezagutza-basearen estaldura txikiagatik erantzun zuzena iragartzea oso zaila izan da. Kasu askotan ez dira ezagutza-basean sintomak lortu eta beraz, oso kasu gutxitan erantzun zuzena iragarri da. Arazo honen aurrean, plangintzan adierazi den moduan, konponbide desberdinak aztertu dira. Hala ere, arazoa ezagutza-basearen estaldura izanik, hobetzeko oso gutxi egitea posible izan da.

Bestalde, aipatutako arazoa eta beste garrantzia gutxiagoko arazoak direla eta, plangintzan desbiderapenak eman dira. Baina plangintzan hainbat lan-pakete paraleloan egitea planifikatu zenez, pakete batean arazoak izateak ez du beste paketeetan aurreratzea eragotzi. Hori dela eta, ez da inoiz proiektuaren bideragarritasuna zalantzan jarri.

### 2.6.2 Plangintzarekiko desbiderapenak

Nahiz eta plangintza hertsiki pentsatu, 2.3. taulan eman diren desbiderapenak ikus daitezke. Desbiderapen nagusiak "Ezagutza-basea sortu" adarrean eman dira, ezagutza-basea sortzeko prozesua pentsatu zena baino luzeago eta zailagoa izan baita. "Emangarrien prestaketa" eta "Kudeaketa" adarretan ere desbiderapenak eman dira, baina lehenengoan aurreikusitako baino denbora gehiago eta bestean, aurreikusitako baino denbora gutxiago behar izan denez, konpentsatu egin da. Hala ere, Gantt diagramarekiko ez da desbiderapen nabarmenik eman.

Lan-paketeak	Estimaturia	Errealia	Diferentzia
<b>Kudeaketa</b>	<b>80</b>	<b>70</b>	<b>-10</b>
Plangintza garatu	30	20	-10
Jarraipena eta kontrola	50	50	0
<b>Artearen egoeraren azterketa</b>	<b>80</b>	<b>105</b>	<b>+25</b>
EB-en egoera aztertu	30	50	+20
EB-tan oinarritutako hurbilpenen azterketa	20	20	0
MCQA hurbilpenen azterketa	30	35	+5
<b>Ezagutza-basea (EB) sortu</b>	<b>100</b>	<b>140</b>	<b>+40</b>
Wikidata	70	85	+15
Estaldura handitu	30	50	+20
<b>EB-an oinarritutako sistema egin</b>	<b>60</b>	<b>80</b>	<b>+20</b>
EB esperimentuak diseinatu	5	5	0
EB esperimentuak egin	30	45	+15
EB esperimentuen analisia	25	30	+5
<b>MCQA sistema egin</b>	<b>80</b>	<b>95</b>	<b>+15</b>
MCQA esperimentuak diseinatu	15	15	0
MCQA esperimentuak egin	50	70	+20
MCQA esperimentuen analisia	15	10	-5
<b>Emangarriak prestatu</b>	<b>75</b>	<b>85</b>	<b>+10</b>
Memoria	60	70	+10
Aurkezpena	15	15	0
<b>Guztira</b>	<b>475</b>	<b>575</b>	<b>+100</b>

2.3 Taula: Lan-pakete bakoitzeko aurreikusitako denbora eta errealitatean erabili den denbora.



## Oinarri teorikoak eta aurrekariak

Adimen Artifiziala (AA), [McCarthy](#)-ren arabera, makina adimenduak, batez ere ordenagailu programa adimenduak, egiteko zientzia eta ingeniari da. Giza adimena ulertzeko antzeko zereginarekin lotuta dago, baina AA ez da naturan behatutako baliabideetara mugatzen. Nahiz eta adimena definitzea oso zaila izan, makina bat gizaki adimentsu baten ataza berdinak egiteko gai bada, adimentsua dela kontsideratzen da. Bi paradigma nagusi daude AA-ean: ezagutzan oinarritutako sistemak eta datuetan oinarritutako sistemak. Ezagutzan oinarritutako sistemetan ezagutza kodifikatu egiten da lengoia formalak erabiliz, eta inferentzia erregelen bidez lengoia horretan idatzitako adierazpenen gainean arazoitu egiten da. Datuetan oinarritutako sistemek, aldiz, datuetatik patroiak erauziz bere ezagutza propioa lortzen dute eta atazak egiten ikasten dituzte adibideetatik abiatuz. Honi ikasketa automatikoa deritza eta Hizkuntzaren Prozesamenduak honen parte bat barnean hartzen du.

Hizkuntzaren Prozesamendua (HP)<sup>1</sup>, *Natural Language Processing (NLP)* ingelesez, gizakion hizkuntza analizatzeko eta adierazteko konputazio tekniken bilduma da. Hala eta guztiz ere, makinenzako hizkuntza ulertzearen ataza zaila izaten da hizkuntzaren ulermen sakona behar izateaz gain, hizkuntza gehienak zabalak eta anbiguoak direlako [10].

HP-ak hizkuntzalaritzarekin erlazio estua duenez, testua eta ahotsa lantzeko (tokenizazioa, adibidez) eta analisi morfologikoa (lematizazioa, adibidez), sintaktikoa (*parsing*, adibidez) eta semantikoa (entitate izendunen ezagutza, *Named Entity Recognition (NER)* ingelesez, adibidez) lantzeko atazak tratatzen ditu, guzti horiek aplikazio desberdinak dituztelarik, hala nola sentimenduen analisia, itzulpen automatikoa, galdera-erantzun sistemak, ahotsaren sintesia, etab.

Oro har, HP-ak giza hizkuntza prozesatzeko, ulertzeko eta harekin elkarri eragiteko aukera ematen die ordenagailuei, eta horrek aplikazio praktiko eta sozial ugari ditu domeinu desberdinetan, medikuntzan adibidez. Medikuntza testuak, hala nola testu klinikoak, txosten medikoak, etab. harturik ataza desberdinak egin daitezke. Izan ere, NER erabili daiteke medikuntzako entitateak testu klinikoetan identifikatzeko, honi ClinicalNER deritza. Gainera, medikuntza testuen sailkapena egin daiteke edota medikuntzako test azterketen

---

<sup>1</sup>Aurreratzean eta sinplifikatzearen, HP laburtzapena erabiliko dugu.

aukera egokia iragarri daiteke galdera-erantzun anitzeko sistemak erabiliz, *Medical Multiple Choice Question Answering* (MedMCQA), deritzona.

Kapitulu honetan, garatutako sistema ulertzeko beharrezkoak diren oinarri teorikoak azalduko dira. Ondoren, HP-ak medikuntza domeinuan izandako bilakaera azalduko da.

## 3.1 Oinarri teorikoak

### 3.1.1 Hurbilpenak

AA-ko problemak ebazteko hurbilpen desberdinak daude. Alde batetik, ezagutza-baseetan oinarritutako hurbilpenak daude, non problemak ebazteko domeinu bateko informazioa duten ezagutza-baseak erabiltzen diren. Beste alde batetik, ikasketa automatikoa erabiltzen dituzten hurbilpenak daude. Horietan algoritmoak eta matematikako modeloak erabiliz, sistemak datuetan patrioiak identifikatzen ikasten du.

#### 3.1.1.1 Ezagutzan oinarritutako metodoak

Lan honetan medikuntzako terminologia eta gaixotasunekin lan egingo da. Hori dela eta, garrantzitsua da, hobeto ulertze aldera, medikuntzarekin erlazioa daukaten ezagutza-baseak ezagutzea, horiek gaixotasunak sailkatzeko eta identifikatzeko erabiltzen direlako. Atal honetan medikuntzako ezagutza-base desberdinak azaltzen dira. Horiek hobeto ulertzeko azalpenaz gain, 3.1. taulan bi gaixotasun desberdinekin (*arthritis* eta *alcoholic liver disease*) ezagutza-base bakoitzean gordetzen den informazioa agertzen da. Horrela, bakoitzarekin lortu daitekeen informazioa eta beraien arteko desberdintasunak ikus daitezke.

### CIE

Clasificación internacional de enfermedades (CIE)<sup>3</sup>, ingelesezko bertsioko *International Statistical Classification of Diseases and Related Health Problems (ICD)* gaztelaniazko lista bat da gaixotasunak, sintomak, aurkikuntza anormalak, kekek, gizarte gorabeherak eta zauri edo gaixotasunen kanpo zergatiak zerrendatzen dituena. CIE bertsiio desberdinak daude eta 2022. urtetik aurrera CIE-11, hau da, 11. bertsiota erabiltzen da. Baina, CIE-10 eta CIE-9 oraindik asko erabiltzen dira. CIE-9-k, CIE-10-ak eta CIE-11-k kode kopuru desberdinak dituzte kapitulu desberdinetan banatuta daudelarik. Zehazki, bakoitzak 14000, 68000 eta 55000 kode desberdin ditu, hurrenez hurren.

Aipatu behar da CIE-10-ak eta CIE-9-k, aldaketak jaso dituztela eta horiek CIE-10-CM eta CIE-9-CM bertsiotan jasotzen direla, non CM "Modificación clínica" esan nahi duen.

### NCI

NCI Thesaurus ID<sup>4</sup> National Cancer Institute (NCI) Thesaurus-en erabiltzen den identifikatzaile bat da. NCI Thesaurus biomedikuntzako ontologia bat da minbizi eta antzeko arloetan termino estandarretarako erabiltzen dena. Minbiziko ikerkuntzarekin zerikusia duten konzeptuak, terminoak, hala nola minbizi motak, botikak, geneak eta beste biomedikuntzako

<sup>3</sup><https://eciemaps.mscbs.gob.es/ecieMaps/browser/metabuscador.html>

<sup>4</sup><https://ncithesaurus.nci.nih.gov/ncitbrowser/>

Ezagutza-basea	Arthritis	Alcoholic liver disease
<b>CIE</b>	<b>CIE-10:</b> <a href="#">M19.90</a> kodea (“Otras artrosis y las no especificadas” M19 kapitulu dena) <b>CIE-9-MC:</b> <a href="#">716.9</a> kodea (“Enfermedades del sistema osteo-mioarticular y tejido conjuntivo” 13. kapitulu dagoena)	<b>CIE-10:</b> <a href="#">K70.10</a> kodea (“Enfermedad alcohólica de hígado” K70 kapitulu dena) <b>CIE-9-MC:</b> <a href="#">571.1</a> kodea (“Enfermedades del aparato digestivo” 9. kapitulu dagoena)
<b>NCI</b>	<a href="#">C2883</a> NCI Thesaurus kodea dauka. Bertan beste NCI terminoekin erlazioak (adibidez, <i>Arthropathy</i> gurasoa, <i>Psoriatic Arthritis</i> semea, <i>Disease_Has_Associated_Anatomic_Site</i> erlazioa dauka <i>Connective and Soft Tissue</i> -rekin...), NCI Metathesaurus <sup>2</sup> <a href="#">C4014799</a> baliokidea, etab. dago gordeta.	<a href="#">C34783</a> NCI Thesaurus kodea dauka. Bertan beste NCI terminoekin erlazioak (adibidez, <i>Non-Neoplastic Liver and Intrahepatic Bile Duct Disorder</i> gurasoa, <i>Alcoholic Cirrhosis</i> semea, <i>Disease_Has_Associated_Anatomic_Site</i> erlazioa dauka <i>Digestive System</i> -ekin...), NCI Metathesaurus <a href="#">C0023896</a> baliokidea, etab. dago gordeta.
<b>MeSH</b>	<a href="#">D001168</a> kodea dauka. MeSH zuhaitza ere agertzen da, non ikus daiteke <i>Arthritis</i> , <i>Experimental</i> semea eta <i>Joint Diseases</i> gurasoa dela, adibidez.	<a href="#">D008108</a> kodea dauka. MeSH zuhaitza ere agertzen da, non ikus daiteke <i>Hepatitis</i> , <i>Alcoholic</i> semea eta <i>Liver Diseases</i> gurasoa dela, adibidez.
<b>SNOMED CT</b>	<a href="#">372091005</a> kodea dauka eta bi deskribapen: “Arthritis (finding)” eta “Arthritis”	<a href="#">41309000</a> kodea eta lau deskribapen: “Alcoholic liver damage (disorder)”, “Alcoholic liver damage”, “ALD - Alcoholic liver disease” eta “Alcoholic liver disease” ditu. Gainera, atributuak, hala nola, non sortzen den, zergatik eta zeintzuk diren bere gurasoak eta bere semeak agertzen dira.
<b>UMLS</b>	<a href="#">C0003864</a> kodea eta “Disease or Syndrome” mota semantikoa dauka. Bestalde, beste hiztegia eta ontologieta-ko definizio desberdinak eta horietan dauden kontzeptuen arteko loturak ditu barnean (adibidez, SNOMED CT, ICD10-CM...).	<a href="#">C0023896</a> kodea eta “Disease or Syndrome” mota semantikoa dauka. Bestalde, beste hiztegia eta ontologieta-ko definizio desberdinak eta horietan dauden kontzeptuen arteko loturak ditu barnean (adibidez, SNOMED CT, ICD10...).
<b>Wikidata</b>	<a href="#">Q170990</a> identifikadore bakarra dauka eta hurrengo <i>statement</i> -ak ditu: Description, Also Known as (arthritic), Symptoms and signs (joint inflammation), Drug or therapy (salsalate...), Health speciality (rheumatology), Anatomical location (joint), kode desberdinak, garrantzitsuenak: DiseasesDB, ICD-9, ICD-10-CM, MedlinePlus ID, MeSH, UMLS, NCI Thesaurus ID, ICPC 2 ID	<a href="#">Q558404</a> identifikadore bakarra dauka eta hurrengo <i>statement</i> -ak ditu: Description, Also Known as (Alcoholic), Symptoms and signs (jaundice...), Possible treatment (preventive medicine...), Health speciality (gastroenterology), Has cause (alcohol consumption), Afflicts (alcoholic), Medical examinations (complete blood count...), Has part or parts (fatty liver disease...) eta kode desberdinak, garrantzitsuenak: ICD-9, ICD-10, ICD-11, MedlinePlus ID, MeSH

**3.1 Taula:** Ezagutza-base desberdinekin gordetzen den informazioa *Arthritis* eta *Alcoholic liver disease* gaixotasun bakoitzarekin.

terminoak eta 50000 erlazio baino gehiago gordetzen ditu. Kontzeptu bakoitzak kode alfanumeriko bakarra du eta modu zehatz batean NCI Thesaurus datu-basean kontzeptuak identifikatzen eta erreferentziatzen laguntzen du.

#### **MeSH**

MeSH<sup>5</sup> (Medical Subject Headings) kontrolatuta eta hierarkikoki antolatuta dagoen hiztegi bat da. Biomedikuntzako artikulua eta osasun arloko literatura indexatu, sailkatu eta bilatzeko erabiltzen da.

MeSH-en dauden terminoak egitura hierarkiko batean modura antolatuta daude, zuhaitz baten antzera. Termino horiek anatomia, gaixotasun, botika, prozedura, genetika eta abarrekin erlazioa dute bakoitzak identifikadore bat (ID MeSH) daukalarik.

#### **SNOMED CT**

SNOMED CT<sup>6</sup> (Systematized Nomenclature of Medicine – Clinical Terms) munduko doitasun, garrantzia eta zabalera handien duen terminologia kliniko integral, eleanitz eta kodifikatua da. Honen helburu nagusia osasun arloan erabiltzen diren esanahiak kodetzea eta datuen erregistro kliniko eraginkorra babestea da, pazientearen arreta hobetzeko helburuarekin. SNOMED CT-k aurkikuntza klinikoak, sintomak, diagnostikoak, prozedurak, gorputz atalak, organismoak, substantziak, botikak eta gailuak gordetzen ditu.

SNOMED CT-k 80000 inguru kontzeptu ditu eta horiek erlazioen bidez lotzen dira beste kontzeptu batzuekin. Bestalde, kontzeptuak hainbat termino kliniko eta esaldien bidez deskribatuak daude, horiei deskribapenak deritze.

#### **UMLS**

UMLS (*Unified Medical Language System*) [11] software eta fitxategi multzoa da osasun eta biomedikuntza hiztegiak, estandarrak eta ontologiak, haien artean, CIE, SNOMED-CT eta MeSH, bateratzen dituenak.

UMLS-k Metathesaurus, sare semantikoa eta SPECIALIST lexikoia ditu bere barnean. Metathesaurus UMLS-ren oinarria da eta kontzeptuen arabera antolatuta dago, kontzeptu bakoitza beste hiztegi batzuetako antzekoak diren izenekin lotuta dago. Gainera, Metathesaurus-ek kontzeptuen arteko erlazioak identifikatzen ditu. Metathesaurus-eko kontzeptu bakoitzari kategoriak (mota semantikoak) esleitzen zaizkie eta elkarren artean lotzen dira erlazio semantikoen bidez, sare semantikoa osatuz. Azkenik, SPECIALIST lexikoaiak ingelesezko hiztegiak eta MEDLINE, UMLS Metathesaurus eta biomedikuntzako terminoak ditu, bakoitzeko informazio sintaktiko, morfologiko eta ortografikoa gordetzen dituelarik.

#### **Wikidata**

Wikidata<sup>7</sup> ezagutza-base libre eta irekia da. Wikidatak Wikimedian dagoen informazioa modu egituratuan gordetzeko biltegi zentral gisa jokatzen du. Arlo askotako informazioa

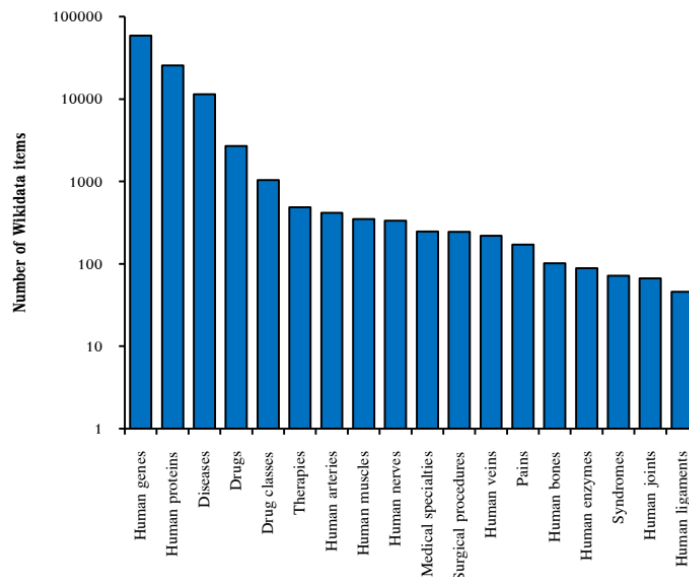
<sup>5</sup><https://www.nlm.nih.gov/mesh/meshhome.html>

<sup>6</sup><https://www.snomed.org/?lang=es>

<sup>7</sup>[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

biltegitzen du, hala nola medikuntza, geografia, historia, etab-eko artikuluetako informazio laburtua gordetzen du RDF<sup>8</sup> tripletetan (entitate1, erlazio, entitate2). Wikidatak *item*-ak gordetzen ditu edozein gai, kontzeptu eta objektu adierazten dutelarik. *Item* bakoitzak identifikadore bakarra dauka eta bakoitzeko dagoen informazioa *statement*-etan gordetzen da. Azkenengo horiek, zehazki, gako-balio bikoteak dira eta horiek propietate bat (identifikadore bat daukana) entitate baten edo batzuen balioekin lotzen dute. Adibidez: "esnea zuria da" esaldiarekin, kolorea propietatea lotuko zen "zuria" balioarekin eta azkenengo hori "esnea" *item*-arekin, RDF tripletan adierazten baldin bada: (zuria, kolorea, esnea).

Wikidata informazio-iturri oso aberats eta erabilgarria da. Adibidez, gaixotasunen inguruan hizkuntza desberdinetan informazio asko gordetzen du; hala nola sintomak, jatorria, diagnostikoak, etab. 3.1. irudian ikus daiteke biomedikuntza domeinuan dagoen kontzeptu kopurua, motaren arabera sailkatuta.



**3.1 Irudia:** Biomedikuntza domeinuan Wikidatan dagoen kontzeptu kopurua, motaren arabera. [Turki et al.](#)-ren artikulutik aterata

### 3.1.1.2 Ikasketa automatikoan oinarritutako metodoak

Ikasketa Automatikoak garapen handia jaso du konputazio ahalmena eta datu kopurua handitu eta algoritmoak hobetu ahala. Honi esker, hasiera batean erabiltzen ziren erregeletan oinarritutako algoritmo sinpleak neurona-sarengatik ordezkatu dira.

Neurona-sare desberdinak sortu dira, tamaina eta topologia desberdinak dituztelarik. Adibidez, sekuentziakin aritzeko neurona-sare errekurrentek (*Recurrent Neuronal Network*<sup>9</sup>, *CNN*) daude. Horiek edozein tamainako sekuentziako sarrera jasota sekuentziako elementu bakoitza banan-banan ezkerretik eskuinera edo *Bidirectional Recurrent Neuronal Network (RNN)*<sup>10</sup> kasuan bezala, bi norabideetan, prozesatzen dute. Hala ere, RNN-ek

<sup>8</sup>[https://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](https://en.wikipedia.org/wiki/Resource_Description_Framework)

<sup>9</sup>[https://en.wikipedia.org/wiki/Recurrent\\_neural\\_network](https://en.wikipedia.org/wiki/Recurrent_neural_network)

<sup>10</sup>[https://en.wikipedia.org/wiki/Bidirectional\\_recurrent\\_neural\\_networks](https://en.wikipedia.org/wiki/Bidirectional_recurrent_neural_networks)

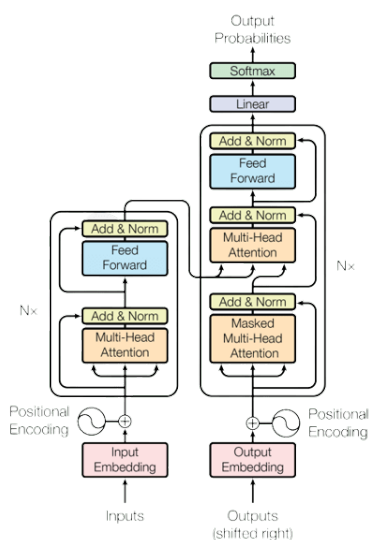
gradienteekin arazoak dituzte eta hori konpontzeko neurona-sarea desberdinak sortu dira adibidez, *Long Short-Term Memory* (LSTM) [12]. Bestalde, neurona-sare konboluzionalak, *Convolutional Neuronal Network* (CNN) [13], sortu dira batez ere argazkiekin aritzeko erabiltzen direnak. Beste neurona-sare desberdinak daude, baina arloa guztiz aldatu zuen Transformer arkitektura da gehien erabiltzen dena gaur egun.

### Transformerrak

Transformer arkitektura [2], aurretik erabiltzen ziren *Recurrent Neuronal Network* (RNN) moduan, datu-sekuentzia ordenatuak kudeatzeko diseinatuta dago. Hala ere, Transformerrek, RNN-ek ez bezala, sekuentzia osoa ez dute modu ordenatuan prozesatu behar, hau da, esaldi baten amaiera prozesatu dezakete honen hasiera aurretik prozesatu gabe. Hori dela eta, Transformerrak paraleliza daitezke konputazio kostuak murriztuz eta gainera, eredu erraldoiak entrenatzea ahalbidetzen dute.

Alde batetik, RNN-ek sekuentzia luzeak tratatzeko erabiltzen direnean arazoak izaten dituzte. Horiek sekuentzien hasierako testuingurua ezin dute erabili, sekuentziak hain luzeak izanik hasierako informazioa galtzen dutelako. Hori konpontzeko, Transformerrek **atentzio** mekanismoa gehitzen dute. Atentzioak sarrerako elementu bakoitzaren garrantzia adierazten du eskalar baten bidez eta gainera, hasierako testuingurua erabiltzea ahalbidetzen du. Beste alde batetik, Transformerrek sekuentziako ordenari ez diote garrantzirik ematen, hura ez baitute sekuentzialki prozesatu behar. Beraz, funtsezkoa da elementu bakoitzak sekuentzian duen posizioaren inguruan informazioa gehitzea. Horretarako **kodeketa posizionala** erabiltzen da, hitzen posizioak eta ordena kodetzeko erabiltzen dena.

Hori guztia horrela izanda, Transformerrek kodetzaile-deskodatzaile arkitektura erabiltzen dute hainbat elementurekin batera, haien artean atentzio-mekanismoa eta kodeketa posizionala dago, 3.2. irudian ikus daitekeenez. Transformerrek arloaren egoera adierazten dute eta HP-ko ataza askotan erabiltzen dira, hala nola itzulpen automatikoan, NER-n, galdera-erantzun sistemetan, etab.



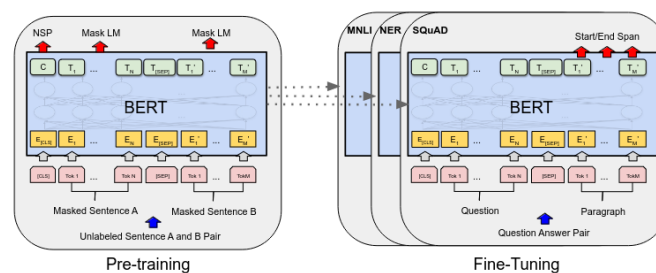
3.2 Irudia: Tranformerraren arkitektura [2]. artikulutik hartua.

### Hizkuntza-eredu neuronalak: BERT

Hizkuntza-eredu bat hitzen sekuentzien gaineko probabilitate-banaketa bat da. m luzerako hitzen edozein sekuentzia emanda, hizkuntza-eredu batek probabilitate bat esleitzen dio sekuentzia osoari:  $P(w_1, \dots, w_m)$ . BERT (Bidirectional Encoder Representations from Transformers) [3] Transformer arkitekturako kodetzailea bakarrik erabiltzen duen eta hortaz, testua sortu ezin duen hizkuntza-eredu bat da. BERT-ek atentzioa bi norabidetan (ezkerrera eta eskuinera) erabiltzen du eta horrela, sekuentziako hitz baten ezker eta eskuin testuingurua kontuan hartzen du.

BERT aurre-entrenatzeko Wikipedia eta BooksCorpus [14] erabili dira. Horiek tokenizatu dira WordPiece tokenizazioa [15] erabiliz ([CLS] tokena erabiliz esaldi hasiera adierazteko eta [SEP] tokena esaldiak bereizteko) eta horrela 30000 token dituen hiztegi bat lortu da. Modu horretan, hiztegi hori hartuz eta bi ataza, Maskaradun Hizkuntza-Eredua (Masked Language Modeling) eta Hurrengo Esaldiaren Iragarpena (Next Sequence Prediction), burutuz aurre-entrenatua izan da. Lehenengo atazan, esaldi bakoitzeko tokenen %15a ausaz ezkutatu da ezkututako token horiek iragartzeko asmoarekin. Horretarako, bi norabideetako atentzioaz baliatu da, eskuineko eta ezkerreko testuingurua eskuratzu. Bigarren atazan, aldiz, esaldien arteko erlazioak ikasteko asmoarekin, bi esaldi emanik horiek erlazio, hau da, elkarren segidan dauden iragarri behar da.

Gauzak horrela, BERT-ek hitzen eta esaldien errepresentazioak eta horien erlazioak jandak ikasiak ditu. Normalean HP-ko atazetan (sentimendu analisia, NER, galdera-erantzun sistemak...) erabiltzen da, baina hobeto aritzeko *fine-tuning* egiten zaio. *Fine-tuning* aurre-entrenatutako modelo bat berriz entrenatzea da ataza konkretuak hobeto eabazteko. Horretarako, egokiak diren parametroak birkalkulatu egiten dira ataza konkretuko datuekin (ikus 3.3. irudia). Aipatu behar da BERT hizkuntza-ereduaz gain, BERT “familiako” beste hizkuntza-eredu ezagunak daudela, ELMo [16], adibidez.



3.3 Irudia: [3]. artikulutik aterata, BERT-aren aurre-entrenamendua eta *fine-tuning*-a.

### 3.1.2 Hizkuntzaren Prozesamenduko atazak

Lan honetan Hizkuntzaren Prozesamenduko bi ataza erabiltzen direnez horiek azalduko dira jarraian.

#### 3.1.2.1 Entitate izendunen ezagutza (NER)

Entitate izendunen ezagutza, *Named Entity Recognition (NER)* ingelesez<sup>11</sup>, informazio erauzketaren azpiatza bat da, zeinetan testuetan agertzen diren entitate izendunak identifikatu

<sup>11</sup>Aurreratzean eta sinplifikatzean, euskaraz ere, NER laburtzapena erabiliko dugu.

eta kategoria zehatz batzuetan sailkatzen diren. Ohiko NER atazetan entitateak pertsona-izenetan, leku-izenetan eta erakunde-izenetan sailkatu ohi dira. Adibidez, Barak Obama-pertsona, EHU-erakundea, Euskadi-lekua.

NER informazioaren kudeaketarako funtsezkoak diren beste arlo askotarako oinarria da, hala nola anotazio semantikoa, galdera-erantzun sistemak, ontologiaren populatzea eta iritzien meatzaritza [17]. Bestalde, NER erabilgarri den lehenengo pausu bat da HP-ko ataza askotan. Adibidez, sentimenduen analisisan nahiko genuke jakin kontsumitzaile baten sentimenduak entitate batengan. Gainera, entitateak erabilgarriak dira testua egituratu-tako Wikipedia bezalako ezagutza-basearekin lotzeko. Era berean, NER funtsezkoa da errepresentazio semantikoak eraikitzeke atazetan, hala nola erlazioak lortu [18].

NER ataza zaila da alde batetik, entitate baten mugak ezartzea, hau da, erabakitzea entitate bat zer den eta zer ez, zaila delako eta beste alde batetik, hizkuntza anbigua delako. NER-en sekuentzia etiketatzeke notazio estandarra *Beginning-Inside-Outside* edo BIO [19] etiketatzea da. Modu honetan, NER erabili daiteke sekuentziak hitzez-hitz etiketatzeke, etiketa bakoitzak entitate bakoitzaren mugak eta mota ezartzen duelarik. BIO etiketatzeak *B* etiketa erabiltzen du entitate baten hasiera adierazteke; *I* etiketa entitate baten barruan dagoela adierazteke eta *O* etiketa entitate kanpokoa dela adierazteke. 3.4. irudian aurkezten da adibide bat, lehenengo esaldia *BIO* etiketatzerik gabe eta bigarrena *BIO* etiketatzearekin, hain zuzen ere.

PER		LOC
Barack Obama	was born in	Honolulu
B-PER	I-PER	O
Barack Obama	was born in	Honolulu

**3.4 Irudia:** NER etiketatzearen adibide bat. Goian *BIO* etiketatzerik gabe, behean *BIO* etiketatzearekin.

### 3.1.2.2 Multiple choice question answering (MCQA)

Galdera-erantzun sistemak, *Question Answering (QA)* ingelesez, gizakiok hizkuntza erabiliz egindako galderak erantzuteke sistemak dira. Horretarako, batzuetan erantzuna webgunean dauden testuetatik edo dokumentu bilduma handietatik ateratzen da, honi *information-retrieval-based question answering* deritzo. Beste batzuetan, aldiz, erantzuna datu-base baten gainean *query*-ak eginez lortzen da, honi *knowledge-based question answering* deritzo [18].

Erantzun anitzeko galdera-erantzun ataza [4], *multiple choice question answering (MCQA)* ingelesez<sup>12</sup>,  $q$  galdera bat, hainbat  $k$  erantzun posible  $O = \{O_1, \dots, O_k\}$  eta kasuaren arabera, aukera bakoitzeko testuinguru bat  $C = \{C_1, \dots, C_k\}$  emanik, erantzun egokia edo egokiak iragartzean datza emandako testuinguruen laguntzaz. QA-en ez bezala, non erantzuna datu-basean edo dokumentuen edo testuen pasarteetan dagoen, MCQA-en erantzunek ez dute zertan testuan egon behar eta batzuetan testu-inferentzia beharrezkoa izaten da. Gainera, MCQA-ak idatzizko ulermen konplexuagoa, eragiketa aritmetikoak, logika eta zentzuzko arrazoiketa, etab. behar du emaitza onak lortzeko. Izan ere, 3.5. irudiari

<sup>12</sup> Aurrerateran eta sinplifikatzearen, euskeraz ere, MCQA laburdura erabiliko duGu.



erreparatzen badiogu, sistemak 1. erantzuna ondo erantzuteko elkarrizketa osoa ulertu behar du eta zentzua erabiliz ondorioztatu behar du elkarrizketa hori bakarrik klase-kideen artean eman daitekeela; "My mother says that, too." esaldia ez lioke inoiz esango arreba batek anai bati. 2. erantzunean, aldiz, ondorioztatu behar du "You'll forget your head if you're not careful." testuko esaldiaren eta erantzuneko "He is too careless." testuaren arteko erlazioa eta hortaz, ondorioztatu hori dela erantzun egokia. Hori guztia gutxi balitz, MCQA ereduak entrenatzeko dauden *dataset* gehienak QA ereduak entrenatzeko daudenak baino txikiagoak dira. Beraz, aipatutako guztiagatik esan daiteke MCQA ataza zailagoa dela.

Dialogue	
W:	Come on, Peter! It's nearly seven.
M:	I'm almost ready.
W:	We'll be late if you don't hurry.
M:	One minute, please. I'm packing my things.
W:	The teachers won't let us in if we are late.
M:	Ok. I'm ready. Oh, I'll have to get my money.
W:	You don't need money when you are having the exam, do you?
M:	Of course not. Ok, let's go... Oh, my god. I've forgot my watch.
W:	<b>You'll forget your head if you're not careful.</b>
M:	My mother says that, too.
<b>Question 1:</b> What's the relationship between the speakers?	
A.	Brother and sister.    B. Mother and son.    C. Classmates. ✓
<b>Question 2:</b> What does the woman think of the man?	
A.	He is very serious.    B. <b>He is too careless.</b> ✓    C. He is very lazy.

**3.5 Irudia:** DREAM *dataset*-eko adibide bat [Jin et al.](#)-ren artikulutik aterata dagoena. ✓: erantzun egokia adierazten du.

MCQA ataza aurrera eramateko, aurre-entrenatutako hizkuntza eredu handiak, BERT adibidez, erabili dira, artearen egoerako emaitzak lortuz. Baina bere potentziala murriztua ikusia izan da MCQA *dataset*-etan dagoen datu faltagatik.

## 3.2 Aurrekariak

Azken urteotan biomedikuntzako literatura kopurua oso azkar hazten ari da. Jadanik dagoen bibliografia kopuru ikaragarriari etengabe gehitzen zaizkio aurkikuntza eta ikuspegi berriei buruzko informazio baliotsua duten txostenek. Adibidez, JAMA (Journal of the American Medical Association) aldizkariak 2022. urtean 13000 artikulua<sup>13</sup> jaso zituen, %18ko gorakada 2021ekin konparatuz, eta PubMed-ek bakarrik 30 milioi artikulua<sup>14</sup> inguru zituen 2021an.

Horren guztiaren aurrean, beharrezkoa da datu biomedikoak errepresentatzeko irudikapen eleaniztun, irakurterraz, editagarri eta egituratua izatea. Azken urteotan, taxonomia handi desberdinak, hala nola MeSH, SNOMED-CT eta UMLS, eta erlazio taxonomikoak ez dituzten ontologiak sortu dira datu biomedikoekin. Nahiz eta ontologiak oso ekarpen baliagarriak izan, horiek beste ontologietan dauden kontzeptuen edota erlazioen falta izaten dituzte. Gainera ontologiak askotan beraien artean bateraezinak dira erabiltzen dituzten sistemengatik. Arazo horiek direla eta, datu biomedikoen integrazioa eta elkarreragingarritasuna zaila izaten da. Bestalde, arazok agerian uzten dute ontologia biomediko handi, bakar, zentralizatu, irakurterraz eta editagarri baten beharra, zeinetan kontzeptu, erlazio eta ontologia guztiak bateragarriak diren. Horregatik, Wikidata sortu da [1].

Nahiz eta HP-an hizkuntza-ereduak asko erabili, Wikidata bezalako ezagutza-baseak oso erabilgarriak izan daitezke. Adibidez, [Liu et al.](#)-ren eta [Marro et al.](#)-ren lanetan sistemen

<sup>13</sup><https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2790524>

<sup>14</sup>[https://www.nlm.nih.gov/bsd/serfile\\_addedinfo.html](https://www.nlm.nih.gov/bsd/serfile_addedinfo.html)

*accuracy* hobetzeko ezagutza-baseak erabili dituzte. Bestalde, Wikidatako datuen izaera egituratua eta lotua aprobetxatuz, [Diomedi and Hogan](#)-ren lanean NER eta erlazio erauzketa egiten da. Wikidata galdera-erantzun sistemetan ere erabil daiteke, batez ere *knowledge-based question answering* hurbilpenean eta zehazki, lan honetan *multiple choice question answering* ataza ebazteko Wikidata erabiliko da.

Hainbeste testu eta ezagutza-base desberdinekin HP-ak lan asko egin dezake honen oinarria testua baita. Historia klinikoen testuen informazioa prozesatuz gero, hainbat helburu medikorekin erabil daiteke, besteak beste; erabaki klinikoak hartzen laguntzea, erakundeen auditoria eta fakturazioa, kalitate klinikoa hobetzea eta ikerketa [23]. Biomedikuntzako testuak hartuz, hitz medikoen desanbiguazioa egin daiteke esanahi desberdinak izan ditzaketen hitz medikoen benetako esanahia lortuz, [Pesaranghader et al.](#)-ren eta [Liu et al.](#)-ren artikuluetan egin den bezala. Baita testu klinikoetan medikuntzako entitateen identifikazioa egin daiteke [Liu et al.](#)-ren lanean bezala, eta gero, hauen arteko erlazioak atera daitezke; adibidez, gaixotasunak eta botikak erlazionatzeko [27], [28]. Bestalde, medikuntzako azterketak hartuz, aurrera sakonean ikusiko den [Pal et al.](#)-ren lanaren antzera, horien erantzun egokia iragartzen saia daiteke medikuntzako ikasleak laguntzeko asmoz. Hala ere, hori guztia gutxi balitz, [Pandey et al.](#)-ren artikuluan beste aplikazio mediko gehiago aurkezten dira.

HP-ko atazak aurrera eramateko, 3.1.1.2 atalean azaldu den bezala, oso ohikoa da BERT bati *fine-tuning* aplikatzea. Baina horiek domeinu orokorreko testuen gainean entrenatuta daude, Wikipedian adibidez. Hortaz, zaila da biomedikuntzako testuetan izango duten portaera aurreikustea, testu medikoen errepresentazio egokia behar da. Gainera, corpus orokorretan eta biomedikuntzako corpusetan hitzen distribuzioa nahiko desberdina denez, domeinu orokorrean entrenatutako hizkuntza-ereduen portaera biomedikuntzako atazetan zalantzarria da. Biomedikuntza bezalako domeinu espezializatuetan, aurretik egindako lanek erakutsi dute domeinu-testua erabiltzeak emaitzak hobetzen dituela. BioBERT [30] izeneko eredu adibidez, BERT bat entrenatzen da PubMed-eko laburpenen eta *PubMed Central*-eko artikuluen osoen (PMC) gainean. PubMedBERT [31], aldiz, PubMed artxibategiko artikuluen laburpenen eta PubMedCentral artikuluen gainean aurre-entrenatuta dago. Bi BERT horien kasuan, biomedikuntzan espezializatua egoteak biomedikuntzako atazetan, NER-en adibidez, emaitzak hobetzen dituela frogatu da. Domeinu klinikoan, aldiz, ClinicalBERT [32] ereduak emaitza onak lortu ditu txosten klinikoekin. Hori dela eta, esan daiteke medikuntza arloan HP atazak burutzea domeinu orokorrean atazak burutzea baino zailago dela. Domeinuan espezializatuak diren testuak eta hizkuntza-ereduak, zeinek aldi berean testu medikoen beharra duten, eta medikuntzan ezagutza behar baita [33].

Lan honetan HP-ko bi ataza egin dira medikuntza domeinuan aplikatuak. Hortaz, horien inguruko arteari buruzko egoera jarraian azalduko da.

#### 3.2.1 NER medikuntza domeinuan

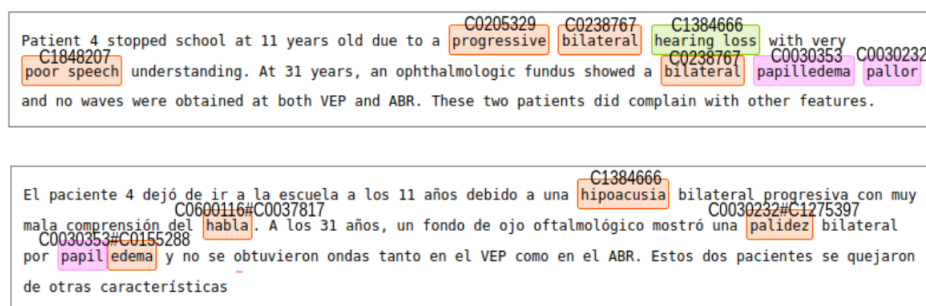
NER ataza literaturan asko landu da. "*Entitate izendun*" terminoa lehenengo aldiz 6th MUC konferentzian [34] erabili zen. Hemendik aurrera NER inguruko interesa piztu da eta ekarpen desberdinak egin dira konferentzia ezberdinetan, adibidez: CoNLL03 [35], ACE [36], IREX [37] eta TREC Entity Track [38]. NER orokorraz haratago, domeinu espezializatuagoetan aritzen diren NER sistemak garatu dira, batez ere medikuntzaren arloan. Medikuntzaren domeinuan NER erabiltzearen helburua, garrantzitsuak diren entitate klinikoak, hala nola

gaixotasunak, sintomak, botikak... detektatzea da, hori guztia, testu klinikoetan. Ataza honi Clinical NER [39] edo NER klinikoa egitea deritzo. Hala ere, Medical Named Entity Recognition (MER) izena hartzen du orokorrean domeinu medikoan egiten den NER-ak.

MER oso erabilgarria izan daiteke ataza desberdinak aurrera eramateko; hala nola sintomak testuan identifikatuta gaixotasunak iragar ditzakegu, edo adibidez, testuko sintomak dagozkion testuko gaixotasunekin lot ditzakegu. Era berean, gaixotasunak eta botikak deskribatzen dituzten entitateak ezagutuz gero, gaixotasunen eta horiek sendatzeko emandako botikak egokiak diren jakiten saia gaitzke, etab [33].

NER klinikoan eta biomedikoan azken urteetan aurrerapen nabarmenak egin dira. Aurrerapen horiek azken urteotan garatutako ELMo eta BERT moduko hizkuntza-ereduak erabilia, sortutako sistemen eskutik etorri dira. Adibidez, ELMo-k emaitza oso onak lortu zituen kontzeptu klinikoak erauzten [40]; beste lan batean, aldiz, BERT erabili zuten kontzeptu klinikoak erauzketa hobetzeko [41].

MER atazan artearen egoerako emaitzak, ingelesezko testuekin, Lee et al.-ren lanean lortu dira %91,22ko F1 balioarekin<sup>15</sup>. Gainera, lan askotan MetaMap [42] eta cTAKES [43] sistemekin, besteak beste, lortu dira F1 balioak hobetzea biek medikuntzako kontzeptuak detektatzen dituztelako. Lan honetan medikuntzako entitateak detektatu dira bai ingelesezko, bai gaztelaniazko medikuntza azterketetan. Ingelesezko entitateak detektatzeko MetaMap erabili da eta gaztelaniaz, aldiz, UNIMER [44] erabili da zeinak %88,45, %89,35 eta %87,57 F1, doitasun eta estaldura balioak lortzen dituen. Lanean erabiliko diren sistemak ulertzeko 3.6. irudian goian ingelesezko esaldi bat eta bertan laranja Metamap-ek, berdez cTAKES-ek eta morez biek detektatuko lituzketen entitate medikoak agertzen dira. Horiekin batera entitate mota eta UMLS kodea bueltatzen dute, besteak beste. Behean, aldiz, gaztelaniaz esaldi bera agertzen da eta bertan UNIMER-ek bueltatutakoa agertzen da, laranja detektatutako sintomak eta morez detektatutako gaixotasunak daudelarik UMLS kodearekin batera.



**3.6 Irudia:** Goiko esaldian laranja Metamap-ek, berdez cTAKES-ek eta morez biek detektatuko lituzketen entitate medikoak, Nixon et al.-ren lanatik ateratako adibidea. Beheko esaldian UNIMER-ek bueltatutako laranja detektatutako sintomak eta morez detektatutako gaixotasunak

### 3.2.2 Medical multiple-choice question answering (MedMCQA)

MCQA ataza azken urteotan ikertua izan den eta oraindik aktibo dagoen arloa da duen konplexutasunagatik. MCQA-aren artearen egoeran teknika desberdinak aplikatu dira erantzun zuzena zein den iragartzeko, horiek bai ikasketa sakona, bai Transformerretan

<sup>15</sup><https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>

oinarritutako eta atentzio mekanismoa erabiltzen duten hizkuntza-ereduak erabiltzen dituzte.

MCQA sistema batzuk sare-neuronaletan oinarritu dira; [Chaturvedi et al.](#)-ren lanean, *Convolutional Neuronal Network*-ak (CNN) [13], [Roy et al.](#)-ren lanean BERT eta CNN hibrido bat eta [Tan et al.](#)-ren lanean, *Long Short-Term Memory*-ak (LSTM) [12] erabili dituzte. Beste sistema batzuetan, arazoa ezagutza-grafo egitura batera hurbildu dute [48], [49]. Lehenengoan, BERT erabili dute galderak eta erantzunak kodetzeko eta gero, grafoaren gainean arrazoitzeko *Graph Neural Network* (GNN) bat [50] erabili dute, modu horretan erantzun zuzena lortuz. Bigarrenak, ez ditu hizkuntza-ereduak erabiltzen, baizik eta ezagutza-grafo bat erabiltzen du, non erantzunak eta galderak erpin moduan eta horien arteko erlazioa ertz moduan errepresentatzen diren; egitura honi esker erantzun zuzena iragartzen du. Bestalde, beste sistemek hizkuntza-ereduak erabiltzen dituzte [4, 51, 52]. Izan ere, hizkuntza-ereduekin artearen egoerako emaitzak lortu dira. Zehazki, [Liu et al.](#)-ren artikuluan aurkeztutako RoBERTa hizkuntza-ereduarekin MCQA-ko *dataset* garrantzitsuetan – SQuAD bi bertsioetan [54] eta RACE-an [55] – artearen egoeraren emaitzak lortu dira: 94,6-ko (v1) eta 89,4-ko (v2) F1 balioa eta 83,2-ko *accuracy* hurrenez hurren.

Nahiz eta aipatu diren hurbilpen guztiak domeinu irekikoak izan, medikuntza domeinuan ere sistemak eraiki dira. Hala ere, literatura gutxi dago MedMCQA-ren inguruan eta oraindik asko falta zaio garatzeko.

MedMCQA (Medical MCQA) arloak daukan lehenengo arazoa datu falta da. Arloak *dataset* konplexuak, domeinuan espezializatuak eta erantzun aukera anizkoitzak dituztenak behar ditu, baina *dataset* handi eta kalitate oneko gutxi daude. Hala ere, badaude ingelesez *dataset* gutxi batzuk [56], gaztelaniaz [8] baina honek badauka ingelesez itzulpena, txineraz [57, 56] eta frantsesez [58]. PubMedQA [59] eta emrQA [60] biomedikuntza domeinuan dauden beste QA *dataset*-ak dira, baina horiek ez dituzte aukera-anitzeko galderak, baizik eta erantzuna testuan aurkitzen da. Bestalde, medikuntzan guztiz espezializatuak ez dauden *dataset*-ak daude, [Hendrycks et al.](#)-ren lanean ingelesezko medikuntzako galderez gain, beste irakasgai batzuen galderak daude, matematika eta humanitateak, adibidez. Azkenik, [Pal et al.](#)-ren lanean MedMCQA *dataset*-a deskribatzen da (4.1.2. atalean sakonean azalduta) eta hori da lan honetan erabiliko den *dataset* bat.

Bestalde, dagoen beste arazo bat ebazteko teknika anitz garatuak ez egotea da. Lan gehienetan BERT bezalako hizkuntza-ereduei *fine-tuning* egin zaie horiek duten jakintzaz baliatuz erantzun zuzena iragartzeko [56, 58, 6, 57]. Beste gutxi batzuetan, bestelako teknikak aplikatzen saiatu dira. [Sileo et al.](#)-ren lanean ezagutza-baseak erabili dituzte okerrak diren erantzunak distrakzio moduan sortzeko eta gero, BERT erabili dute erantzun zuzena iragartzeko. [Liu et al.](#)-ren lanean ere ezagutza-baseetan dagoen informazioa aprobetxatzen dute erantzun zuzena iragartzeko. Azkenengo honek, gainera, aukeratu duen erantzuna arrazoitzen du. Izan ere, Medical MCQA arloa apurka-apurka hazten den bitartean garrantzitsua da modeloak aukeratu duen erantzuna arrazoitzeko gai izatea, horrela, ikusi ahal izateko zer ikasi duen modeloak. Baina oso lan gutxi daude honen inguruan, agerian utziz arloaren heldutasun eza.

Aipatu den guttia kontuan hartuz, arloan lan asko dago egiteke. Lan honetan [Pal et al.](#)-ren lana hartu da erreferentzia moduan eta bertan aurkeztzen den MedMCQA *dataset*-a, lau aukera bakarrik dituena, hartu da eta dagozkion aldaketak egin zaizkio bost aukera dituen beste *dataset* baten gainean probatu ahal izateko.

# Garatutako inplementazioa

Kapitulu honetan garatutako inplementazioa azaltzen da. Lehenik eta behin, erabiliko diren datuak azalduko dira 4.1 azpiatalean. Atal horretan sortutako medikuntzarako baliabide aberats eta eleaniztuna ere azalduko da erabilitako datuak direlako. Ondoren (4.2 azpiatala), erabiliko diren metrikak azalduko dira eta, amaitzeko, problema ebazteko erabili diren bi hurbilpenak azalduko dira, 4.3 eta 4.4 azpiataletan.

## 4.1 Datuak

Lan honetan hiru datu-multzo desberdin erabili dira iturri desberdinetatik aterata daudenak. Atal honetan horiek azalduko dira bakoitzak daukan informazioa deskribatuz.

### 4.1.1 CasiMédicos-MIR

Egindako bi hurbilpenen – ezagutza-baseetan oinarritutako hurbilpena eta MCQA hurbilpena – zehaztasuna neurtzeko CasiMédicos-MIR *dataset*-a erabiliko da. CasiMédicos-MIR IXA taldeak sortutako *dataset* bat da MIR azterketekin. Azterketak CasiMédicos<sup>1</sup> orritik aterata daude galdera bakoitzak erantzun zuzena eta arrazoia dituelarik. Aipatu behar da erantzun egokiak eta zergatiak mediku bolondresen idatzi dituztela eta horren ondorioz, kasu batzuetan okerrak izan daitezkeela.

Corpusa MIR urte desberdinen 622 galdera klinikoz osatuta dago. MIR azterketa Espainian egiten den azterketa bat izanda, galderak gaztelaniaz bakarrik daude, baina ingelesera itzulpena ere egin da. Gauzak horrela, bi corpus sortu dira, bata gaztelaniaz eta bestea ingelesez eta lan honetan biak erabiliko dira. Zehazki, ingelesezko eta gaztelaniazko corpusak ezagutza-baseetan oinarritutako hurbilpenean erabiliko dira eta MCQA hurbilpenean, aldiz, ingelesezkoa bakarrik erabiliko da. Hala ere, aipatu behar da corpusek galderak errepikatuak zituztela eta, beraz, horiek ezabatu ostean 508 galdera dituen bi corpusekin lan egin dela.

Corpuseko galderak erantzun-anizkoitzekoak dira, bakoitzak 4 edo 5 erantzun posible ditu, bakarra zuzena izanik. Corpus bakoitza *csv* fitxategi batean dago dagokion hizkuntzaren arabera.

<sup>1</sup><https://www.casimedicos.com/mir-2-0/>

#### 4. GARATUTAKO INPLEMENTAZIOA

tzan galdera bakoitzeko hurrengo informazioa gordetzen delarik: urtea, galdera horren id konkretua, galdera horren id orokorra, galdera horren amaiera id-a, baliogabetu den ala ez, galdera, erantzun zuzenaren zergatia, erantzun zuzenaren zergatia azaltzeko erabilitako erreferentziak, galdera mota (neurologia, dermatologia..), aukera 1, aukera 2, aukera 3, aukera 4, aukera 5 eta aukera egokiaren zenbakia (1etik 5era). Aipatu behar da 4 erantzun posible dituzten galderetan “aukera 5” hutsik dagoela. 259 kasu daude 4 erantzun posiblerekin eta 5 erantzun posiblerekin, aldiz, 249 kasu daude. Kasu bakoitzak gordetzen duena grafikoki 4.1. taulan ikus daiteke.

year	question_id _specific	full_question	full_answer	type	option_1	option_2	option_3	option_4	option_5	correct _answer
2011	40	Un hombre presenta se forma brusca astenia e ictericia detectándose unas transaminasas superiores a 2000 UI/L. Los marcadores serológicos muestran el siguiente patrón: Anti VHA IgM negativo, HBsAg negativo, AntiHbc IgM positivo, anti VHC negativo. ¿Cuál es el diagnóstico?	Una interesante de algo que en general hay que estudiarse. Por la forma de presentación, es claramente una hepatitis aguda con lo que la 1 se descarta. La IgM de VHA negativa descarta la 5 y el HBs Ag negativo la 4 (para tener virus delta necesitas antígeno de superficie B). Al tener IgM anti Hbc hablamos de una hepatitis aguda B (3), ya que hay un periodo ventana de negativización del antígeno.	DIGESTIVO	Hepatitis crónica B.	Hepatitis aguda no viral.	Hepatitis aguda B.	Sobreinfección por virus D (delta).	Hepatitis aguda A y B.	3
2022	119	Mujer de 13 años, sin antecedentes relevantes, con menarquia hace 3 meses, seguida desde los 10 años por escoliosis idiopática que ha empeorado. En la exploración física presenta una giba de 7 grados en el test de Adams y en el escoliograma una curva toracolumbar T4-L1 de 35 grados de Cobb y un Risser 0. La actitud correcta a tomar será:	Ortesis. Angulo de Cobb entre 25° - 45°. Esqueleto inmaduro (Risser 0).	TRAUMATOLOGÍA	Recomendar natación y revisión en tres meses.	Prescribir una ortesis tipo corsé.	Derivar a fisioterapia para elasticación del raquis.	Revisar en 6 meses con una nueva radiografía.		2

**4.1 Taula:** CasiMédicos-MIR *dataset*-ean gaztelaniaz dauden bi kasuren adibidea. Lehenengoak 5 aukera ditu eta bigarrenak, aldiz, 4 bakarrik. Hutsik zeuden zutabeak ezabatu dira.

#### 4.1.2 MedMCQA *dataset*-a

MedMCQA *dataset* gutxi egon arren Pal et al.-ren lanean aurkeztzen dena erabiliko da. Bertan *train*-a Indiako medikuntza ikasketetan sartu ahal izateko simulakro eta *online* azterketen 182.822 galderaz osatuta dago, *test*-a 1991. urtetik aurrera All India Institute of Medical Sciences (AIIMS PG) aukera anitzeko azterketen 6.150 galderaz osatuta dago eta *dev*-a 2001. urtetik aurrera National Eligibility cum Entrance Test (NEET PG) azterketetako 4.183 galderaz osatuta dago.

Kasu/galdera bakoitzeko 4 erantzun posible eta galdera motaren arabera erantzun zuzen bakarra edo anitzak zeintzuk diren adierazita dago, zehazki bakoitzeko: id, galdera, aukera 1, aukera 2, aukera 3, aukera 4, aukera egokia, galdera mota (erantzun bakar edo anitz ondo), azalpena, irakasgai eta gaia gordetzen da. Azaldutakoa 4.2. taulan ikus daiteke.

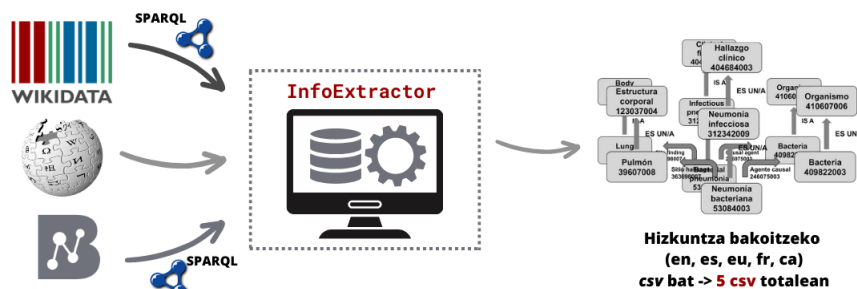
id	question	opa	opb	opc	opd	cop	choice_type	exp	subject_name	topic_name
b64a9cd7-d076-4c55-8be1-f9e44fcec6cc	Which of the following is the only complete cartilage ring in the respiratory tree?	Cricoid cartilage	Thyroid cartilage	Cuneiform cartilage	Epiglottis	0	single	Complete cartilage ring is seen in Cricoid cartilage. Cricoid cartilage is shaped like a ring. It encircles the larynx below the thyroid cartilage. It is thicker and stronger than the thyroid cartilage.	Anatomy	

**4.2 Taula:** MedMCQA *dataset*-aren adibide bat.

### 4.1.3 Sortutako ezagutza-basea

Wikipedia informazio iturri oso aberatsa da. Honek gaixotasunen inguruan informazio asko gordetzen du, hala nola sintomak, kausak, diagnostikoak, CIE kodeak, Snomed kodeak... Kode horiek gaixotasunen arteko erlazioak ezartzen dituzten ontologiaren parte dira; adibidez, erlazio hierarkikoak (CIE-9 eta CIE-10), parte-hartze erlazioak (UMLS), etab. Gainera, informazio guzti hori hizkuntza desberdinetan gordetzen du, informazio hori desberdina izanik lengoaia batetik bestera. Wikipedian, adibidez, *Fibrilación Ventricular* artikulua *infobox*-ean, katalanez, gaztelaniaz ez dauden sintomak daude.

Hori horrela izanik, hasiera batean gizakion gaixotasunak zituen ezagutza-base bakar bat egitea pentsatu zen bost hizkuntza desberdinetan (ingeleza, frantsesa, euskara, katalana eta gaztelania) osagarri zegoen informazioa bateratuz. Baina azkenean, 4.1. irudian ikus daitekeenez, erabaki da hizkuntza bakoitzeko bana egitea eta Wikidata, Wikipedia eta BabelNet erabiltzea informazioa lortzeko, Wikidata izanik iturri nagusia. Bertan hizkuntza desberdinetan dagoen informazioa berdina da, beste hizkuntzen itzulpena egiten baita. Hori dela eta, hizkuntza batetik bestera informazio osagarririk ez dago.



4.1 Irudia: Ezagutza-basea sortzeko egin den sistema.

Atal honetan gaixotasunen ezagutza-base eleaniztun bat sortzeko egin den sistema azalduko da. Horretarako, Wikipediako informazioa jaisteko ezagutza-base desberdinen azterketa bat egin da, hasteko. Beraz, lehenik eta behin, erabilitako baliabideen azterketa azalduko da. Ondoren, ezagutza-basea nola sortu den eta prozesuan topatutako arazo nagusiak azalduko dira. Azkenik, lortutako ezagutza-basea aurkeztuko da honek gordetzen duen informazioa deskribatuz.

#### 4.1.3.1 Aztertutako baliabideak

Helburua Wikipediako informazioa jaistea denez, hasiera batean Wikipediatik zuzenean informazioa jaistea pentsatu zen. Wikipediak artikulua askotan *infobox*-etan laburtzen du artikuluen informazioa, 4.2. irudian, gorri *arthritis* gaixotasunaren ingelesezko artikulua *infobox*-ak gordetzen duen informazio baliagarria ikus daiteke. Hori dela eta, *infobox*-etatik informazioa jaistea pentsatu zen.

Horretarako, Python-eko *BeautifulSoup* eta *wptools* liburutegiak erabiltzea pentsatu zen horiek *web scraping*<sup>2</sup> egiteko eta *infobox*-ak jaisteko pentsatuta daudelako, hurrenez hurren. Baina azkenean metodo hori ez zen erabili hizkuntza guztietan informazioa jaisteko algoritmo bat definitzea zaila zelako. Izan ere, hizkuntzaren arabera informazioa, *infobox*-ak adibidez, desberdin gordetzen da; horiek ingelesez *box*, gaztelaniaz *Ficha*, katalanez

<sup>2</sup>[https://en.wikipedia.org/wiki/Web\\_scraping](https://en.wikipedia.org/wiki/Web_scraping)

## 4. GARATUTAKO INPLEMENTAZIOA

**Arthritis** is a term often used to mean any disorder that affects *joints*.<sup>[4]</sup> Symptoms generally include *joint pain* and stiffness.<sup>[2]</sup> Other symptoms may include redness, warmth, *swelling*, and decreased *range of motion* of the affected joints.<sup>[2][3]</sup> In some types of arthritis, other organs are also affected.<sup>[6]</sup> Onset can be gradual or sudden.<sup>[5]</sup> There are over 100 types of arthritis.<sup>[9][4][5]</sup> The most common forms are *osteoarthritis* (degenerative joint disease) and *rheumatoid arthritis*.<sup>[6]</sup> Osteoarthritis usually occurs with age and affects the fingers, knees, and hips.<sup>[6][10]</sup> Rheumatoid arthritis is an *autoimmune disorder* that often affects the hands and feet.<sup>[6]</sup> Other types include *gout*, *lupus*, *fibromyalgia*, and *septic arthritis*.<sup>[6][11]</sup> They are all types of *rheumatic disease*.<sup>[2]</sup> Treatment may include resting the joint and alternating between applying ice and heat.<sup>[6][12]</sup> Weight loss and exercise may also be useful.<sup>[6][13]</sup> Recommended medications may depend on the form of arthritis.<sup>[14][8]</sup> These may include pain medications such as *ibuprofen* and *paracetamol* (acetaminophen).<sup>[8]</sup> In some circumstances, a *joint replacement* may be useful.<sup>[6]</sup> Osteoarthritis affects more than 3.8% of people, while rheumatoid arthritis affects about 0.24% of people.<sup>[15]</sup> Gout affects about 1-2% of the *Western* population at some point in their lives.<sup>[16]</sup> In Australia about 15% of people are affected by arthritis,<sup>[17]</sup> while in the United States more than 20% have a type of arthritis.<sup>[11][18]</sup> Overall the disease becomes more common with age.<sup>[11]</sup> Arthritis is a common reason that people miss work and can result in a decreased quality of life.<sup>[8]</sup> The term is derived from *arthr-* (meaning 'joint') and *-itis* (meaning 'inflammation').<sup>[19][20]</sup>

**Classification**  [ [edit](#) ]

There are several diseases where joint pain is primary, and is considered the main feature. Generally when a person has "arthritis" it means that they have one of these diseases, which include:

....



### 4.2 Irudia: Wikipediako ingelesezko *arthritis* gaixotasunaren *infobox*-a.

*Infotaula*, frantsesez *Taxobox* eta euskeraz *infobox* moduan gordetzen dira. Gainera, hizkuntza berdinean informazioa ere ez da beti berdin gordetzen; katalanez *Infotaula* moduan *infobox*-ak bilatuz gero, beti ez da informazio guztia lortzen.

Beste alde batetik, Wikipedia informazio iturri oso aberatsa izanda honen informazioa gordetzen duten ezagutza-base desberdinak daude, horien artean DBPedia. Hori dela eta, gaixotasunak eta horien inguruko informazioa jaisteko DBPedia<sup>3</sup> erabiltzea pentsatu zen. Baina ingelesez gaixotasunak lortu ostean baztertu zen DBPedia zaharkitua dagoelako. Izan ere, informazio gehien daukan hizkuntza ingelesa izanik, 2464 gaixotasun baino ez ziren lortu. Gainera, 4.3. taulan ikus daitekeenez, beste ezagutza-baseetan informazio gehiago gordetzeaz gain, interesatzen zaigun informazio (4.3. taulan beltzez) gehiago dago beste iturri batzuetan.

#### 4.1.3.2 Sorkuntza

Ezagutza-basea sortzeko egin denaren adierazpen grafikoa 4.1. irudian ikus daiteke. Bertan adierazten den bezala, Wikidata, Wikipedia eta BabelNet erabili dira informazioa lortzeko, Wikidata eta BabelNet kasuetan SPARQL kontsultak erabili dira. Gauzak horrela, InfoExtractor sistema bidez gure ezagutza-basea osatuko duten 5 csv-ak lortu dira. Atal honetan sakonean azalduko da jarraitutako prozesua.

Aipatu diren iturriak baztertu ostean, gaixotasunak eta horien inguruko informazioa erauzteko ezagutza-base nagusia Wikidata izatea erabaki da. Alde batetik, 4.3. taulan ikus daitekeenez, gaixotasunen inguruan interesatzen zaigun informazio gehien gordeta daukan iturria delako eta beste alde batetik, biomedikuntzarekin erlazioa duten 102226 kontzeptu (gaixotasunak, botikak, geneak...) dituelako [1].

Inplementazioa azaldu baino lehen, alde batetik jakin behar da Wikidatatik informazioa erauzteko SPARQL<sup>4</sup> kontsultak erabiltzen direla eta, beste alde batetik, Wikidatan kontzeptu guztiak kode baten bidez identifikatuak daudela.

<sup>3</sup><https://www.dbpedia.org/>

<sup>4</sup><https://en.wikipedia.org/wiki/SPARQL>



Gaixotasuna	Wikidatako propietateak	DBPediako propietateak	Wikipedia infobox-etako propietateak
Arthritis	<ul style="list-style-type: none"> <li>• <b>Symptoms and signs</b></li> <li>• <b>Drug or therapy</b></li> <li>• <b>Health speciality</b></li> <li>• <b>DiseasesDB</b></li> <li>• <b>ICD-9</b></li> <li>• <b>ICD-10-CM</b></li> <li>• <b>MedlinePlus ID</b></li> <li>• <b>MeSH</b></li> <li>• <b>UMLS</b></li> <li>• <b>NCI Thesaurus ID</b></li> <li>• Anatomical location</li> <li>• ICPC 2 ID</li> <li>• Description</li> <li>• Also known as</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Symptom</b></li> <li>• <b>Treatment</b></li> <li>• <b>Medication</b></li> <li>• <b>Type</b></li> <li>• <b>DiseasesDB</b></li> <li>• <b>ICD-9</b></li> <li>• <b>ICD-10</b></li> <li>• <b>Medlineplus</b></li> <li>• <b>meshId</b></li> <li>• <b>SameAs</b></li> <li>• <b>ICD</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Symptoms</b></li> <li>• <b>Treatment</b></li> <li>• <b>Medication</b></li> <li>• <b>Specialty</b></li> </ul>
Alcoholic liver disease	<ul style="list-style-type: none"> <li>• <b>Health speciality</b></li> <li>• <b>ICD-9</b></li> <li>• <b>ICD-10</b></li> <li>• <b>MedlinePlus ID</b></li> <li>• <b>MeSH</b></li> <li>• <b>ICD-11</b></li> <li>• <b>Symptoms and signs</b></li> <li>• <b>Has cause</b></li> <li>• <b>Possible treatment</b></li> <li>• Afflicts</li> <li>• Medical examinations</li> <li>• Has part or parts</li> <li>• Description</li> <li>• Also known as</li> </ul>	<ul style="list-style-type: none"> <li>• -</li> <li>• <b>ICD-9</b></li> <li>• <b>ICD-10</b></li> <li>• <b>Medlineplus</b></li> <li>• <b>meshId</b></li> <li>• <b>Synonyms</b></li> <li>• <b>SameAs</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Specialty</b></li> </ul>

**4.3 Taula:** Wikidata, DBPedia eta Wikipedia arteko konparazioa. *Arthritis* eta *Alcoholic liver disease* gaixotasun bakoitzarekin lor daitezkeen propietateak. Horiek parekatuta daude iturri desberdinen beste propietateekin eta beltzez interesatzen zaizkigun propietateak daude.

Gauzak horrela, eman den lehenengo pausoa Wikidatako gaixotasun guztien kodeak lortzea da SPARQL kontsulta baten bidez. Kontsulta hori, egingo diren kontsultak nolakoak diren ikusteko, jarraian aurkezten da. Bertan i tem2 guztiak lortzen dira, horiek gaixotasun klaseko (wd:Q12136) azpiklasekoak diren i tem guztiak izango dira. Baina gizakion gaixotasunez osatutako ezagutza-base bat sortzea helburua izanik, kontsulta horrek itzultzen dituen kode guztiak ez dira interesgarriak, itzultitako kodeen artean animalien eta landareen gaixotasunak, eskualdeen izenak, greben izenak, etab. baitaude. Hori dela eta, beste kontsulta baten bidez lortu nahi ez diren hurrengo kontzeptuen kodeak lortu dira: animalien eta landareen gaixotasunak, landareen nahasmendu fisiologikoak, siamdarrak, izurriteak, parafiliak, gose-greba, pozoitzeak, metal pozoitzeak, garraioak eta fetitxismoa. Azkenik, hasierako kodeetatik bigarren kontsulta bidez lortutako kodeak ezabatu dira, gizakion gaixotasun kodeekin bakarrik geratuz.

```
SELECT ?item2 ?item2Label WHERE {
  ?item (wdt:P279*) wd:Q12136. # wdt:P279* = azpiklase guztiak
  ?item2 (wdt:P31) ?item. # wdt:P31 = instantzia da
  SERVICE wikibase:label { bd:serviceParam wikibase:language
    ↪ "[AUTO_LANGUAGE],en". } # Helps get the label in your language,
    ↪ if not, then en language
}
```

Behin eskuratu nahi ditugun kontzeptuen kodeak ditugula, kode bakoitzeko, hizkuntza bakoitzeko (ingelesez, gaztelaniaz, katalanez, frantsesez eta euskaraz) hurrengo egin da:

1. csv fitxategi bat sortu da zeinetan gaixotasun guztiak dagozkien informazioarekin gordeko diren eta beste fitxategi bat sortu da, zeinetan prozesuan sortu diren errore guztiak hots, gorde ezin izan diren gaixotasunen kodeak gordetzen diren.
2. Kode bakoitzeko SPARQL beste kontsulta bat egin da hurrengo informazioa lortzeko. Beltzez aurrerago erabilgarria izango den informazioa dago: **gaixotasunaren izena**, gaixotasun kodea, sintomak, sintomen kodeak, tratamenduak, tratamenduen kodeak, “Different from” propietatea, “Different from” propietatearen kodea, arriskuak, arriskuen kodeak, ICD-9, ICD-10, UMLS, MESH, NCI Thesaurus ID, **Wikipediako artikulua**ren esteka, deskribapena eta **sinonimoak**.
3. Kontsulta egiterakoan erroreren bat sortu bada, 10 saiakera desberdin egin dira eta behin eta berriz errore bat eskuratzen bada, errorearen fitxategian gorde da.

Lortzen diren informazioen artean, asko erabiltzen den NCI Thesaurus ID kodea dago. Hori dela eta, ezagutza-basean dagoen informazioa handitzeko erabili da. Aipatu bezala, NCI-k 500000 lotura inguru ezarrira ditu. Lotura horien artean “Disease\_May\_Have\_Finding” eta “Disease\_Has\_Finding” erlazio motak aurkitzen dira, ingelesezko kontzeptu bat beste baten sintoma dela adierazten dutenak.

Hori guztia kontuan hartuta, ingelesez bakarrik lortu ahal izan dira gaixotasunen sintomak. Beraz, hasteko, gure ingelesezko csv-a hartu da eta gaixotasun bakoitzeko NCI Thesaurus ID kodea eskuratu da, egotekotan. Jarraitzeko, kode horrekin NCI Thesaurus ID webgunean bilaketa bat simulatu da gaixotasun horren informazio guztia lortuz. Informazioa atal desberdinetan banatuta dago, interesatzen zaizkigun loturen informazioa *Relationships* atalean gordeta daudelarik. Beraz, *web scrapping* bidez, *Relationships* atal horretan, gaixotasun kode horrekin, “Disease\_May\_Have\_Finding” eta “Disease\_Has\_Finding” erlazio motak dituzten kontzeptu guztiak lortu dira eta bilatu den gaixotasun horren sintoma moduan erantsi dira csv-an.

Horrela, hizkuntza bakoitzeko gaixotasunak dituzten csv-ak lortu dira. Baina csv-etan oraindik interesgarriak ez diren kodeak daudenez, eskuz gaixotasunak berriro filtratu dira.

Gaixotasun asko lortu arren, estaldura handitzeko asmoarekin beste bi hurbilpen egin dira; batean Wikipedia erabiliz eta bestean BabelNet erabiliz. Aurretik beltzez jarritako propietateak erabilgarriak izango dira hurbilpen horietan.

### Wikipedia bidez estaldura handitu

Wikidatak Wikipediako informazioa gorde arren, ez du Wikipediako informazio guztia. Hori dela eta, Wikipediak gordeta dituen gaixotasun guztiak konparatu dira Wikidata erabiliz, azaldu den prozesua jarraituz, lortu direnekin. Konkretuki, Wikipedian dauden gaixotasunen estekak Wikidatatik lortu diren gaixotasunen estekekin konparatu dira. Horretarako, lehenik eta behin Wikipediako gaixotasunen estekak lortu dira hurrengo pausoak emanez:

1. Wikipedia lista honetako ([https://en.wikipedia.org/wiki/Category:Lists\\_of\\_diseases](https://en.wikipedia.org/wiki/Category:Lists_of_diseases)) gaixotasun guztien estekak lortu dira *web scraping* eginez *Beautiful-Soup* liburutegiarekin. Kontuan hartu da alde batetik, lista horretako gaixotasun batzuen estekak ez direla existitzen eta bestetik, gaixotasun batzuen estekak beste gaixotasun batzuen esteketara birbideratzen direla. Adibidez, AIDS gaixotasunaren esteka HIV/AIDS gaixotasunaren orrira birbideratzen da.
2. Esteka batzuk beste batzuen esteketara birbideratzen direnez, esteka errepikatuta daude. Hori dela eta, errepikatuta daudenak ezabatu dira.

Ondoren, horiek Wikidatatik lortutako gaixotasunekin konparatu dira. Horretarako, Wikipediatik lortutako esteka bakoitza, gure ezagutza-baseko gaixotasun bakoitzak duen estekarekin konparatu da. Esteka berdina ez bada, estekako gaixotasunaren izena Wikidatatiko gaixotasun bakoitzaren izen edota sinonimoekin konparatu da. Laburbilduz, esteka bakoitza gure *csv*-an dagoen begiratu da eta horrela ez bada, estekaren gaixotasunaren izena *csv*-an bilatu da. Modu honetan, Wikipediatik falta diren gaixotasunak lortu dira, 517 gaixotasun, zehazki. Azkenik, falta diren gaixotasun horiek, izenaren bidez, Wikidatan bilatu eta bilaketaren lehenengo emaitza, emaitza hori artikulua edo entsegu kliniko bat ez bada, gure *csv*-tan gehitu dira.

### BabelNet bidez estaldura handitu

BabelNet<sup>5</sup>, alde batetik, entziklopedia hiztegi eleaniztun bat da zeinak terminoen estaldura lexikografiko eta entziklopedikoa duen, eta, beste alde batetik, ontologia bat da zeinak kontzeptuak eta entitateak erlazio semantikoko sare handi batean konektatzen dituen.

BabelNet-ek informazioa, besteak beste, Wikidatatik lortzen du, hori dela eta, estaldura handitzeko erabili da. BabelNet-ek SKOS<sup>6</sup> erabiltzen du era estandar batean datu lexikoak errepresentatu eta antolatzeko. Hortaz, kontzeptu bakoitzak identifikadore bakar bat dauka eta horiek haien artean erlazionatzen dira *skos:broader*, *skos:narrower* eta *skos:related* erlazioen bidez, besteak beste. Lehenengo biek, erlazio hierarkiko bat adierazten dute, *broader* erabiltzen da kontzeptu bat beste baten gainean dagoela adierazteko eta *narrower*, aldiz, kontrakoa adierazteko. Bestalde, *related* erlazioak elkartze-erlazio bat adierazten du, hau da, maila berean dauden erlazioak adierazteko balio du.

Hori jakinda, eman den lehenengo pausua BabelNet-en dauden gaixotasun guztien kodeak lortzea izan da. Horretako, *skos:narrower* propietatea erabili da hierarkian beheago dauden erlazioak interesatzen zaizkigulako. Hots, gaixotasun kontzeptua hartzen baldin

<sup>5</sup><https://babelnet.org/>

<sup>6</sup><https://www.w3.org/TR/skos-reference/>

badugu ez zaizkigu maila berean (sinonimoak) dauden kontzeptuak interesatzen, baizik eta horren azpitik dauden eta erlazioa duten kontzeptuak interesatzen zaizkigu. Horrek esan nahi duelako bietatik gaixotasun kontzeptua orokorra dela eta ondorioz, gaixotasun mota desberdinak lortuko direla (azpitik daudenak). Adibidez, *arthritis* gaixotasunak *broader* erlazioa dauka *inflammatory disease*-ekin eta *narrower* erlazioa *rheumatoid arthritis*-ekin. Beraz, gaixotasun kontzeptuaren kodea hartu da eta horrekin *narrower* erlazioa daukaten gaixotasunen kodeak lortu dira. Kode horiekin erlazio berdina daukaten gaixotasunak lortu dira eta horrela, errekurtsiboki, prozesua errepikatu da gaixotasun berririk ez lortu arte.

Egindako lehenengo pausoarekin gaixotasunen kodeak baino ez dira lortzen. Beraz, kode bakoitzeko BabelNet-en bilaketa bat simulatu da eta *web scraping* bidez lortu dira gaixotasunen izenak.

Nahiz eta BabelNet-ek Wikipediatik informazioa lortu, honek gordeta dituen gaixotasunak Wikidatako gordetakoak baino gehiago diren edo eraikitako ezagutza-basean ez dauden gaixotasun osagarri asko badituen jakitea garrantzitsua da. Izan ere, BabelNet-eko gaixotasun gehienak gure ezagutza-basean baldin badaude, ez du merezi falta direnak Wikidatatik lortzea. Beraz, gaixotasun guztien izenak lortu ostean, gure ezagutza-basean momentu horretan dauden gaixotasunekin konparatu da jarraitzeak merezi duen ala ez jakiteko. Baina, 4.4 taulan ikus daitekeenez, BabelNet-etik lortutako ia gaixotasunen erdia, horien artean, *brain tumor*, *inguinal hernia*, *rheumatism*... adibidez, ez daude eraikitako ezagutza-basean. Hori dela eta, jarraitzea erabaki da. Hala ere, aipatu behar da taulako zenbakiak ikusiz BabelNet-ek Wikidatako baino gaixotasun gutxiago dituela.

BabelNet-etik lortutako gaixotasun artean gure ezagutza-basean dauden kopurua	5454
BabelNet-etik lortutako gaixotasun artean gure ezagutza-basean <b>ez</b> dauden kopurua	9089

**4.4 Taula:** BabelNet-etik lortutako gaixotasunen inguruko emaitzak.

Ondoren, gure ezagutza-basean ez dauden BabelNet-eko gaixotasunen izena lortu eta, Wikipediarekin bezala, izenak Wikidatan bilatu eta gure *csv*-etan gehitu dira. Azkenik, badaezpada ere, berriro filtratu dira gaixotasunak interesatzen ez zaizkigun kodeak ez daudela konprobatu edota ezabatzeko.

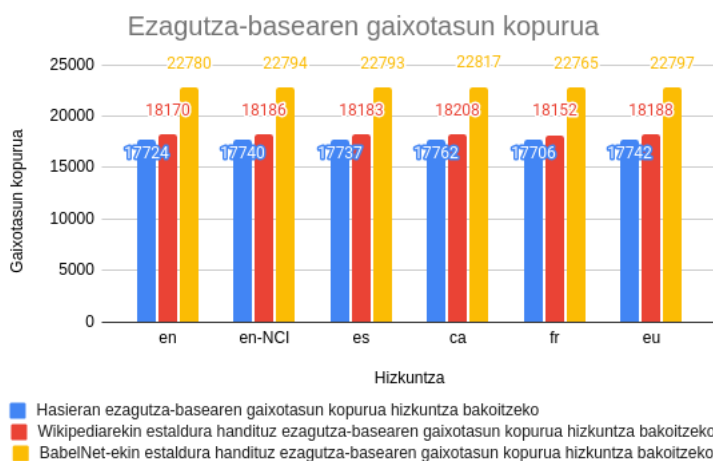
#### 4.1.3.3 Lortutako ezagutza-basearen deskribapena

Aipatutako prozesu luzearen ostean, ezagutza-base handi eta eleaniztun bat lortu da, GitHub-en<sup>7</sup> aurkitu daitekeena, eta HP-ko ataza desberdinak egin edota horien inplementazioa erraztu dezakeena. Emaizta moduan, ingelesez, gaztelaniaz, katalanez, frantsesez eta euskaraz bakoitzeko *csv* bana lortu da zeinetan gaixotasun bakoitzeko izena, gaixotasun kodea, sintomak, sintomen kodeak, tratamenduak, tratamenduen kodeak, “Different from” propietatea, “Different from” propietatearen kodea, arriskuak, arriskuen kodeak, ICD-9, ICD-10, UMLS, MESH, NCI Thesaurus ID, Wikipediako artikularen esteka, deskripzioa eta sinonimoak gordetzen diren. Gainera ingelesez, beste *csv* bat lortu da NCI kodearen bidez lortutako sintoma gehiagorekin, ingelesa bihurtuz gure ezagutza-basean sintoma gehien dituen hizkuntzan.

<sup>7</sup><https://github.com/paulaonta/InfoExtractor>

Bestalde, 4.3. irudian ikus daiteke hasieran ezagutza-baseak zituen gaixotasunak eta Wikipediarekin eta BabelNet-ekin estaldura handitu ostean lortutako gaixotasun kopurua. Ikus daiteke nola bi kasuetan estaldura handitzea lortu den. Hala ere, Wikipedia kasuan momentu horretan ezagutza-basean ez zeuden 517 gaixotasun lortu dira eta horietatik 446 bakarrik aurkitu dira Wikidatan, 446tan bakarrik handituz estaldura. BabelNet-en kasuan, aldiz, 9089 gaixotasun lortu dira eta horietatik 5978 bakarrik aurkitu dira Wikidatan. Alabaina, irudian ikus daiteke BabelNet-en bidez estaldura ez dela 5978tan handitu. Hori da, esan bezala, filtraketa bat egin delako prozesuaren amaieran kode desegokiak kentzeko.

Laburbilduz, medikuntza baliabide aberats eta eleanitza (5 hizkuntzetan) lortu da. Honek 22.000 kontzeptu baino gehiago ditu, bakoitzak gainera, beste ezagutza-baseekin (MeSH, ICD-9, ICD-10, UMLS, NCI, Wikipedia) erlazioa dute. Azkenik, lortutako gaixotasun kopurua ikusirik aipatu behar da 3.1.1.1. atalean agertzen den 3.1. irudian adierazitako gaixotasun kopuru handiagoa lortu dela. Izan ere, ateratako artikulua 2019. urtekoa da eta ezagutza-basea 2022. urtean egin denez, 3 urte horietan gaixotasun berri asko sartu ahal izan dira.



**4.3 Irudia:** Hasieran ezagutza-baseak zituen gaixotasunak eta Wikipediarekin eta BabelNet-ekin estaldura handitu ostean lortutako gaixotasun kopurua hizkuntza bakoitzeko.

## 4.2 Metrikak

Metrikak eredu ebaluatzeko erabiltzen dira eta gure ereduak egin behar duen ataza zein ondo egiten ari diren neurtzeko balio dute. Gainera, aukera ematen digute beste eredu batzuekin konparatzeko.

Atal honetan sistema bakoitza ebaluatzeko arloan erabiltzen diren metrikak eta gure sisteman erabili direnak azalduko dira.

### 4.2.1 Ezagutza-baseetan oinarritutako hurbilpeneko metrikak

Sistema honetan erabiliko den neurria **zehaztasuna**<sup>8</sup>, *accuracy* ingelesez, da. Zehaztasuna sailkapen ereduak ebaluatzeko metrika bat da. Kasu honetan ataza sailkapen ataza baten moduan ikus daiteke non ereduak 1 eta 5 artean sailkatu behar duen galdera bakoitza, zenbaki hori adieraziz erantzun zuzenaren zenbakia. Esan daiteke zehaztasunak ondo sailkatu dituen iragarpenen portzentaia adierazten duela. Honela definitzen da zehaztasuna sailkapen bitarra ez denean:

$$\text{Zehaztasuna} = \frac{\text{Ondo iragarritako kasu kopurua}}{\text{Iragarritako kasu kopuru totala}}$$

Gure sisteman metrika hori erabiltzea erabaki da alde batetik, beste hurbilpenarekin konparatu ahal izateko eta beste alde batetik, sailkapen atazak *dataset* orekatuetan egitean hots, klase guztiak nahiko berdin banatuta daudenean, asko erabiltzen den metrika bat delako. Izan ere, CasiMédicos-MIR *dataset*-a orekatua dagoela esan daiteke; 1 klaseko 97, 2 klaseko 104, 3 klaseko 138, 4 klaseko 111 eta 5 klaseko 58 kasu baitaude. Nahiz eta 5. klasea duten kasuak besteekin konparatuz gutxiago izan, orekatua dagoela esan daiteke kasu guztiek 5 erantzun posible ez dituztelako.

### 4.2.2 MCQA hurbilpeneko metrikak

MCQA ataza sailkapen ataza baten moduan ikus daiteke. Hori dela eta, arloko lanetan [56, 8, 59] aurreko atalean azaldu den **zehaztasuna** metrika erabiltzen ohi da. Gainera, medikuntzako azterketa ofizialetan metrika hori ere erabiltzen da. Bestalde, aurreko sisteman eta erreferentzia moduan hartu den Pal et al.-ren lanean metrika hori erabiltzen denez, bertan lortutako emaitzekin konparatu ahal izango da. Honengatik guztiagatik zehaztasun metrika erabiliko da.

Beste MCQA lanetan [58], **Exact Match Ratio (EMR)** metrika erabiltzen da. Metrika hori erantzun anitz egokiak direnean, hau da, etiketa anitzeko sailkapen ataza denean erabili izaten da. EMR-k batzen ditu guztiz egokiak diren galderak (erantzun posible guztiak egokiak direnean) eta dauden galdera kopuru guztiarekin zatitzen du. Neurriak ez ditu kontuan hartzen partzialki zuzenak diren galderak. Eredua ebaluatzeko ez da neurri hau erabili etiketa anitzeko sailkapen bat egiten ez delako, galdera guztietan erantzun bakarra delako zuzena. Gainera, hori horrela izanda, esan daiteke kasu honetan EMR eta zehaztasuna baliokideak direla EMR-k neurtuko lukeelako zuzen dagoen klase bakarrarekin bat egiten duten kasuak.

## 4.3 Ezagutza-baseetan oinarritutako hurbilpena

Ezagutza-baseak erabilgarriak izan daitezkeela komentatu da aurreko kapitulu batean (3. kapitulu). Hori dela eta, ezagutza-baseetan oinarritutako sistema bat eraikiko da MIR azterketetako galderen erantzun egokia iragartzeko. Atal honetan erantzun egokia iragartzeko sistema azalduko da.

---

<sup>8</sup><https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=es-419>

### 4.3.1 Funtzionalitateak eta hauen inplementazioa

Sistemak CasiMédicos-MIR corpuseko galderak eta erantzun posibleak izanda erantzun zuzena asmatu behar du. Horretarako, hasteko, galdera eta erantzunen NER egingo da bakoitzeko medikuntza entitate nagusiak identifikatuz. Jarraitzeko, galdera bakoitzeko erantzun bakoitzean identifikatu diren gaixotasunak, eraikitako ezagutza-basean bilatuko dira horien sintomak lortuz. Sintomak izanda, amaitzeko, erantzun zuzena iragarriko da. Hots, galderetan dauden sintomekin komunean sintoma gehien dituen erantzuna ontzat hartuko da. Adibide bat 4.4. irudian agertzen da.

Una mujer de 32 años ha viajado a Cuba de donde ha llegado hace tres días. Al día siguiente de su regreso acude a su médico de cabecera por **fiebre** elevada, artromialgias intensas y **cefalea** que le había comenzado antes de iniciar el viaje de regreso y le prescribe paracetamol. ¿Cuál es el diagnóstico más probable?

1. **Fiebre tifoidea.** ✓
2. **Dengue.**

#### SINTOMAK:

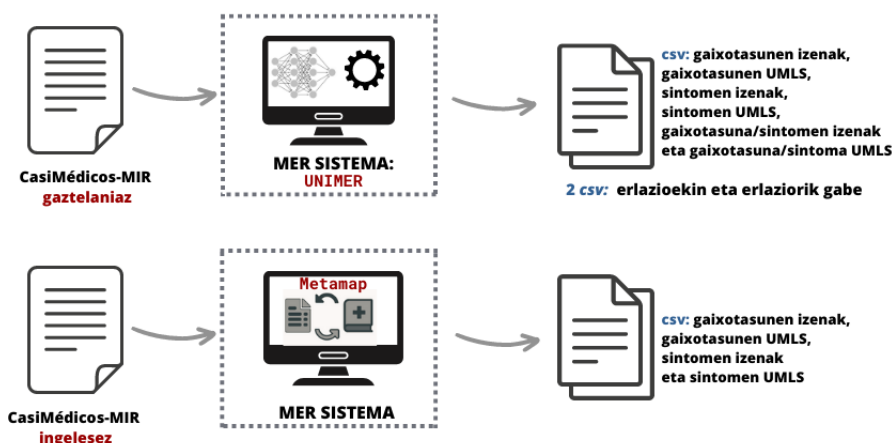
1. Esplenomegalia, bradicardia, cansancio, **cefalea**, delirio, diarrea, distensión abdominal, eflorescencia, estreñimiento, **fiebre**, hemorragia gastrointestinal, hepatomegalia, insomnio, leucopenia, palidez
2. adenopatía, adinamia, anorexia, artralgia, bradicardia, cansancio, **cefalea**, erupción maculopapular, hemorragia, mialgia, náusea, vómito

**4.4 Irudia:** Adibide bat, galderan urdinez dauden sintomak eta gorritz, erantzunetan detektatutako gaixotasunak daude. Erantzunen azpian gaixotasun horiek eraikitako ezagutza-basean bilatuz, dituzten sintomak daude. Beltzez komunean dituzten sintomak daude, lehenengoak gehiago dituenaz, hura zuzena da.

#### 4.3.1.1 Gaixotasun eta sintomen identifikazioa

Lehenik eta behin, galdera eta erantzun posibleetatik entitate mediko nagusiak (gaixotasunak eta sintomak) identifikatuko dira, hau da, MER egingo da.

Corpusa bi hizkuntza desberdinetan dago hortaz, ezin da sistema berdina erabili bietan. Sistemek hizkuntza prozesatzen dutenez eta hizkuntzak desberdinak direnez, hizkuntza baterako ikasten dituen entitateak, kasu honetan medikuntza entitateak, beste hizkuntza batean ez dira berdinak izango ez direlako berdin idazten. Hori horrela, ingelesezko sistema eraikitzeko MetaMap erabiliko da eta gaztelaniazko sistema eraikitzeko UNIMER. Sistemak zer egingo duen eta hizkuntza bakoitzeko zer bueltatuko duen 4.5. irudian ikus daiteke.



4.5 Irudia: Gaixotasunak eta sintomak identifikatzeko sistema.

MetaMap-ek detektatutako entitateak erabiliko dira, horretarako Python-en *script*<sup>9</sup> bat sortuko da dagozkion aldaketak eginez. Honela, galdera eta erantzun bakoitza pasa eta horietan dauden entitateak (errepikatu gabe) itzuliko ditu UMLS kodearekin batera; entitate horietatik bakarrik gaixotasunak eta sintomak, bere kodearekin batera, lortuko dira eta dagokion formatuan gordeko dira: *csv* batean non gaixotasunak, gaixotasunen UMLS kodeak, sintomak eta sintomen UMLS kodeak daudelarik. Adibidez, sistemak hurrengo galderarekin gorritz dauden gaixotasunak eta urdinez dauden sintomak detektatuko lituzke:

*A 27-year-old woman under study in the dermatology department for recurrent episodes of oral aphthous ulcers. She was referred to the gastrointestinal department for a blood test showing Hb 11.5 gr/dl Fe 38 AST 52 ALT 64 Ac antitransglutaminase IgA 177. The patient reported only occasional diffuse abdominal discomfort. She reports no signs of hemorrhage or jaundice. The physical examination is normal. Which of the following tests do you consider to be the most appropriate to make the diagnosis?*

UNIMER-rek, aldiz, galdera eta erantzun bakoitza pasata entitate guztiak bueltatzen ditu tokenak BIO moduan etiketatuta daudelarik. Token bakoitzeko entitate mota (*Alergia, Grp\_Enfermedad, Grp\_Medicamento...*), UMLS kodea eta kodeak esleitzen dion mota semantikoa (*hallazgo, transtorno, organismo...*) itzultzen du, besteak beste. Entitate motak ez du gaixotasuna eta sintoma artean desberdintzen, kode motak, aldiz, bai.

Hori horrela izanik, kode mota begiratu, lehenik eta behin, sintomak (*hallazgo*) eta gaixotasunak (*transtorno, anomalía morfológica, transtorno clínico*) detektatuko dira. Ondoren, kode motak gaixotasuna edo sintoma moduan detektatu ez dituen kasuetan, entitate motan arreta jarriko da, zehazki, *Grp\_Enfermedad* den begiratu da. Nahiz eta entitate horiek kode motak gaixotasun edo sintoma bat moduan ez identifikatu, honek itzultitako UMLS kodea lortuko da (egotekotan). Entitate motak ez duenez sintoma edo gaixotasunak desberdintzen bi modutan (gaixotasuna/sintoma) gordeko dira. Informazio guztia lortu dela, hurrengo formatuan gordeko da: *csv* batean non gaixotasunak, gaixotasunen UMLS kodeak, sintomak, sintomen UMLS kodeak, gaixotasuna/sintoma eta gaixotasuna/sintoma UMLS kodeak daudelarik. Adibidez, sistemak hurrengo galderarekin gorritz dauden gaixotasunak eta urdinez dauden sintomak, aurreko adibide berdinarekin, detektatuko lituzke:

*Mujer de 27 años en estudio en el Servicio de Dermatología por presentar episodios de aftas bucales recidivantes. Es remitida a las consultas de Aparato Digestivo por objetivar en una analítica: Hb 11.5 g/dl, Hto. 35%, Fe 38 ug/dl, AST 52 U/L. ALT 64 U/L, Ac antitransglutaminasa IgH 177 U/ml. La paciente sólo refiere presentar molestias abdominales difusas de forma ocasional. No refiere diarrea, signos de hemorragia ni ictericia. La exploración física es normal. ¿Qué prueba de las que se mencionan a continuación considera que es la más adecuada para el realizar el diagnóstico?*

Ikus daitekeenez, gaztelaniaz sintomak eta gaixotasunak ez ditu guztiz identifikatzen, adibidez, “molestias” sintoma moduan identifikatu du eta “abdominales” kalifikatzaile moduan, baina ez du “molestias abdominales”, elkarrekin, sintoma moduan identifikatzen. Ingelesean, aldiz, biak elkarrekin, “abdominal discomfort”, identifikatu ditu sintoma moduan. Honek arazo bat suposatzen du gaztelaniaz gaixotasun eta sintomen inguruan informazioa galtzen delako.

Informazio hori lortzeko, gaztelaniaz prozesu berdina errepikatuko da baina UNIMER-relations aukerarekin exekutatu. Aukera honen bidez tokenen arteko erlazioak lortuko

---

<sup>9</sup><https://gweissman.github.io/post/using-metamap-with-python-to-access-the-umls-metathesaurus-a-quick-start-guide/>



dira. Erlazio hauek dira *Modif*, hau da, token bat beste bat aldatzen, kalifikatzen duenean, *Causada\_por* token bat beste baten ondorioa denean eta *LOC* token bat beste token baten ondoan kokatuta dagoenean. Gainera, entitate batek hitz bat baino gehiago duenean *MultiE* erlazioa erabiltzen da. Hori jakinda, gaixotasunak eta sintomak eta errekurtsiboki hauek dituzten erlazioak lortzen joango dira, eta informazio guztia aurreko formatu berdinean gordeko da. Modu honetan, gaztelaniaz bi *csv* lortuko dira, bata erlazioekin eta bestea erlazioarik gabe.

Gauzak horrela, 4.5. taulako emaitzak lortu dira. Taulan ikus daitezke ingelesezko eta gaztelaniazko kasu bakoitzeko, galdera eta erantzunekin, lortutako gaixotasun eta sintomen **kode** kopuru desberdinak.

UNIMER-ek detektatzen duen entitate bakoitzarentzat beti ez du kode bakarra identifikatzen; MetaMap-ek, aldiz, beti kode bakarra itzultzen du entitate bakoitzarentzat. Gainera, gaztelaniazko adibidean ikus daitekeenez, entitate batzuk bai gaixotasun bai sintoma moduan detektatzen ditu, hemorragia adibidez. Honengatik guztiagatik, alde batetik 4.5. taulan ikus daiteke, gaztelaniaz kode gehiago detektatu direla, hori baita agerikoa da jarritako adibidea ikusiz. Beste alde batetik, taulan 3. errenkadan UNIMER-ek entitate mota begiratzat detektatu dituen entitateen kode kopuruak agertzen dira, esan bezala, ez du desberdintzen gaixotasuna edo sintoma den eta entitate bakoitzeko beti ez du kodea lortzen.

Bestetik, bi hizkuntzetan galderetan sintoma gehiago daude gaixotasunak baino, eta erantzunetan aldrebes gertatzen da. Hori zen espero zena, izan ere, eraikitako sistemak hori kontuan izanik diseinatu eta eraiki baita.

Azkenik, 4.6. taulan ikus daiteke komunean eta komunean ez dituzten kode kopurua, ez dituzte asko komunean baina hori da gaztelaniaz kode askoz gehiago daudelako. Hala ere, 4.5. taulan ikus daiteke ingelesez 641 gaixotasun kode daudela galderetan eta 4.6. taulan komunean 305 daudela, ia erdia. Gainera, jarritako adibidea aztertuz detektatuko entitateen artean erdia dute komunean.

Amaitzeko, aipatu behar da bai gaztelaniaz bai ingelesez, erantzunetan, galderetan ez bezala, kasu guztietan entitateren bat (gaixotasun edo sintoma) ez duela detektatzen (ikus 4.7. taula). Honek arazo bat suposatzen du 0 erantzunetan entitateak detektatu dituzten kasuetan ez duelako ezer iragarriko edota erantzun bakarrean entitateak detektatu dituzten kasuetan hori izango delako iragarriko duena. Beraz, kasu horiek ezin dira ontzat hartu. Hori dela eta, zehaztasuna kalkulatzera kontuan hartuko da.

	Ingelesez		Gaztelaniaz	
	Galderak	Erantzunak	Galderak	Erantzunak
Gaixotasunak	641	780	1453	1482
Sintomak	615	75	2059	462
Gaixotasunak/sintomak	-	-	301	256 (221 koderik gabe)

**4.5 Taula:** MetaMap-etik eta UNIMER-retik ingelesez eta gaztelaniazko galderak eta erantzunak pasa ondoren lortutako gaixotasun eta sintoma **kode** desberdinak.

#### 4.3.1.2 Gaixotasunak eraikitako ezagutza-basean bilatu

Hizkuntza bakoitzeko sintomak, gaixotasunak eta kodeak lortu eta egituratu direla, erantzunetakoak eraikitako ezagutza-basean bilatuko dira. 4.6. irudian ikus daitekeenez, alde

#### 4. GARATUTAKO INPLEMENTAZIOA

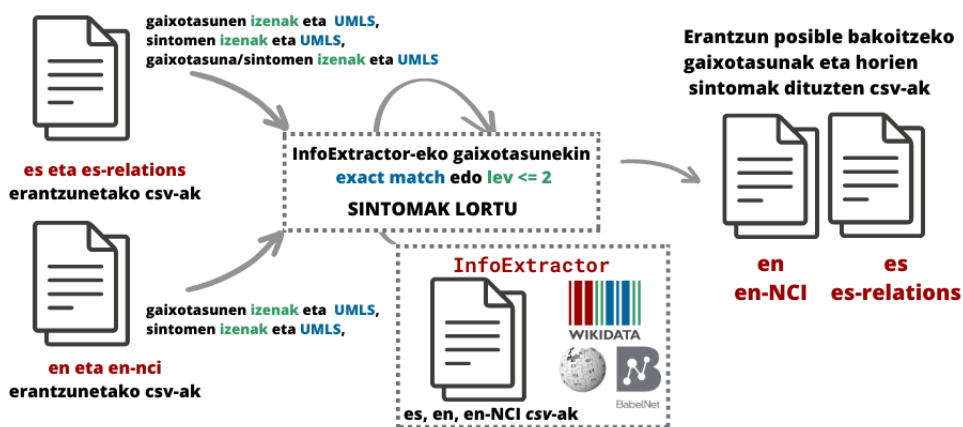
	Komunean		Komunean ez	
	Galderak	Erantzunak	Galderak	Erantzunak
Gaixotasunak	305	412	1484	1438
Sintomak	331	30	2012	477
Gaixotasunak/sintomak	5	12	1547	1087

**4.6 Taula:** MetaMap-etik eta UNIMER-retik ingelesez eta gaztelaniazko galderak eta erantzunak pasa ondoren komunean eta komunean ez dituzten gaixotasun eta sintoma kodeak.

	Zenbat erantzunetan entitateak detektatu					
	0	1	2	3	4	5
<b>Ingelesez</b>	241	76	44	51	63	33
<b>Gaztelaniaz</b>	139	62	37	51	118	101

**4.7 Taula:** Ingelesez eta gaztelaniaz zenbat kasu daude zeinetan entitateak detektatu dituzten erantzun kopurua 0, 1, 2, 3, 4 eta 5 den, hurrenez hurren.

batetik, saiatuko gara UMLS kodeak zuzenean ezagutza-basean bilatzen. Beste alde batetik, kodea ezagutza-basean aurkitu ez diren entitateen izenak eta kodea ez duten entitateen izenak (4.5. taulako 3. errenkadako 221 entitateen izenak), Levenshtein distantzia bidez bilatuko dira ezagutza-basean.



**4.6 Irudia:** Gaixotasunak eraikitako ezagutza-basean bilatzeko sistema.

Levenshtein distantzia [63], zenbaki bat da<sup>10</sup> hitz batetik abiatuta beste hitz bat lortzeko egin behar diren gutxieneko eragiketen kopurua adierazten duena. Erantzunetan detektatu diren entitateen kodeak ezagutza-basean aurkitu ez badira, entitate horien izenen eta ezagutza-baseko gaixotasun guztien izenen arteko distantzia neurtuko da; distantzia 2 baino txikiago edo berdina baldin bada gaixotasun horren sintomak lortuz. Adibidez, erantzun batean C0010346 kodea daukan gaixotasuna ez da ezagutza-basean aurkitu, baina Levenshtein bidez, horren *Crohn disease* izena ezagutza-basean aurkitu da. Zehazki, ezagutza-baseko *Crohn's disease* gaixotasunaren sintomak lortu dira. Bestalde, erantzunetan kodea ez duten entitateen izenen eta ezagutza-baseko gaixotasun guztien izenen arteko

<sup>10</sup>Adibidez, *tonsillitis* Wikidatan agertzen den hitz zuzena balitz eta detektatutako entitatea *tonsillitis*, bien arteko distantzia 1 da *l* bat txertatuz gero Wikidatakoa lortuko genukeelako.

distantzia ere neurtuko da; distantzia 2 baino txikiago edo berdina baldin bada gaixotasun horren sintomak lortuz.

Lehenik eta behin, erantzunetan aurkitu diren gaixotasunen kodeak bilatuko dira 3.1.1.1. atalean azaldu den ezagutza-basean, 4.8. taulako emaitzak lortuz. Badaezpada ere, sintomen kodeak gaixotasun moduan ere bilatuko dira. Hala ere, honek ez du eragin handia izango emaitzetan, erantzunetan sintoma gutxi detektatu direlako eta sintomak gaixotasun moduan bilatuz, seguruenik ez direlako aurkituko.

Taulan, ikus daitekeenez, alde batetik, ingelesez aurkitu diren kode gehiago daude aurkitu ez direnak baino; gaztelaniaz, aldiz, kontrakoa gertatzen da. Gaztelaniaz entitate batzuetan hainbat kode (baina izen bakarra) detektatzen ditu eta kode guztiak ez dira gaixotasunenak. Horregatik ere, gaztelaniaz aurkitu ez diren kode kopurua izen kopurua baino handiagoa da. Gainera, gaztelaniaz erlazioak lortu direnean, izen ez aurkituen kopurua handiago da gaztelaniaz erlazioak kontuan hartu gabe lortutakoak baino. Hori horrela da erlazioak lortzerakoan izen kopurua handitzen delako. Adibidez, “molestias abdominales”, bi izenekin gordetzen da. Alde batetik, “molestias” eta beste alde batetik, erlazioa daukan beste izenarekin: “molestias abdominales”. Hala ere, bi izen horiek izango dituzten kodeak berdinak dira.

Beste alde batetik, bi hizkuntzetan aurkitu diren kode guztien erdia baino gehiagotan ez dira sintomak aurkitu, hau da, Wikidatan ez ditu sintomak gaixotasun horrek. Gainera, ingelesez, gaztelaniaz baino kode gutxiago aurkitu arren aurkitutako sintomak gehiago dira. Honek esan nahi du, eraikitako ezagutza-baseak ingelesez sintoma gehiago dituela.

	Ingelesez	Gaztelaniaz
Kode aurkituak	570	1003
Kode <b>ez</b> aurkituak	285	1197
Izen <b>ez</b> aurkituak	285	802 (1381 erlazioekin)
Kode aurkituen sintomak totalen	915	901
Sintomak ez dituzten kode aurkituak	386	807
Sintomak dituzten kode aurkituak	184	196

**4.8 Taula:** Eraikitako ezagutza-basean aurkitu diren erantzunetako gaixotasun kodeak eta lortutako sintomak ingelesez eta gaztelaniaz.

Ondoren, Levenshtein distantzia aplikatuko da bai UNIMER-rek gaixotasun edo sintoma moduan identifikatu ezin izan dituen eta kodea lortu ezin izan den gaixotasunen izenekin (4.5. taulako 3. errenkadako 221 entitateen izena), bai aurreko pausoaren ostean aurkitu ez diren entitateen izenekin (4.8. taulako 3. errenkada). Horrela, 4.9. taulako emaitzak lortu dira. Taulan ikus daiteke aurkitu **ez** diren izenak gehiago direla. Bestetik, aurkitutako izenetatik erdiak baino gehiagok ez du sintomarik. Gainera, gaztelaniaz erlazioak lortzeari esker izen gehiago eskuratu direnez, sintoma gehiago lortu dira.

Hala ere, esan bezala, eraikitako ezagutza-baseak ingelesez NCI-tik ateratako sintoma gehiago dituen beste csv bat du. Beraz, sintoma gutxi aurkitu direnez, handitzeko asmoz, prozesua errepikatuko da baina ingelesez NCI-tik ateratako sintoma gehiago dituen csv-an begiraturaz. Horrela, 4.10. taulan ikus daitekeenez, sintomak dituzten aurkitutako kode gehiago daude. Gainera, 4.11. taulan ikus daiteke izen eta sintomak dituzten izen gehiago ere aurkitu direla. Bestetik, bi tauletan aurkitutako sintomak handitu dira, NCI sintomak begiratzen direnez, sintoma gehiago lortu dira, aurkitutako izen bakoitzeko Wikidatatik

#### 4. GARATUTAKO INPLEMENTAZIOA

lortutako sintomak eta NCI sintomak lortzen direlako.

	Ingelese	Gaztelaniaz
Izen aurkituak	65	206 + <u>53</u> = 259 565 + <u>106</u> = 671
Izen <b>ez</b> aurkituak	220	596 + <u>168</u> = 764 816 + <u>283</u> = 1099
Izen aurkituen sintomak totalen	296	398 + <u>473</u> = 871 787 + <u>931</u> = 1718
Sintomak ez dituzten izen aurkitu kopurua	39	143 + <u>19</u> = 162 509 + <u>67</u> = 576
Sintomak dituzten izen aurkitu kopurua	26	63 + <u>34</u> = 97 56 + <u>39</u> = 95

**4.9 Taula:** Eraikitako ezagutza-basean Levenshtein distantzia bidez aurkitutako erantzunetako gaixotasunen izenak eta lortutako sintomak ingelesez eta gaztelaniaz. Azpimarratuta UNIMER-rek gaixotasun edo sintoma moduan identifikatu eta kodea lortu ezin izan den gaixotasun entitateen izenekin lortutako emaitzak. Gaztelaniaz bigarren errenkadan **erlazioekin** lortutakoa emaitzak daude.

	Ingelese	Ingelese-NCI
Kode aurkituak	570	570
Kode <b>ez</b> aurkituak	285	285
Izen <b>ez</b> aurkituak	285	285
Sintomak totalen	915	<b>2110</b>
Sintomak ez dituzten kode aurkitu kopurua	386	331
Sintomak dituzten kode aurkitu kopurua	184	<b>239</b>

**4.10 Taula:** Eraikitako ezagutza-basean aurkitu diren erantzunetako gaixotasun kodeak eta lortutako sintomak ingelesez eta ingelesez NCI bidez lortutako sintomekin eraikitako csv-tan bilatuta lortutakoak.

	Ingelese	Ingelese-NCI
Izen aurkituak	65	<b>67</b>
Izen <b>ez</b> aurkituak	220	218
Sintomak totalen	296	<b>606</b>
Sintomak ez dituzten izen aurkitu kopurua	39	36
Sintomak dituzten izen aurkitu kopurua	26	<b>31</b>

**4.11 Taula:** Eraikitako ezagutza-basean Levenshtein distantzia bidez aurkitutako erantzunetako gaixotasunen izenak eta lortutako sintomak ingelesez eta ingelesez NCI bidez lortutako sintomekin eraikitako csv-tan bilatuta lortutakoak.

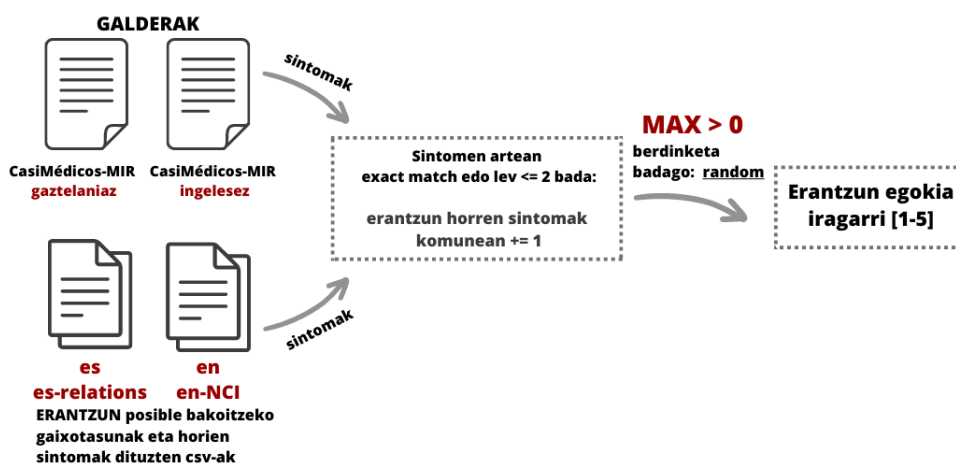
#### 4.3.1.3 Aukera egokia iragarri

Aukera egokia iragartzeko, kasu bakoitzeko erantzunetako gaixotasunetatik lortutako Wikidatako sintomak hartu eta kasu horretako galderak dituen sintomekin konparatuko dira, komunean dituzten sintomak eskuratuz. Prozesu hori galdera bakoitzeko erantzun posible guztiekin errepikatuko da eta komunean gehien daukan erantzuna ontzat hartuko

da. Berdinketa baldin badago ausaz bat aukeratuko da. Horrek algoritmoari ausazkotasuna ematen dionez, *seed* desberdinekin probatuko da: 42, 60, 50, 46 eta 50.

Bestalde, erantzun batek galderarekin komunean sintomarik ez baldin badu ez da egokia izan daitekeen aukera moduan kontsideratuko, hau da, adibidez, 5 erantzunek sintomak lortu baldin badituzte ezagutza-basean eta guztiek galderan dauden sintomekin komunean sintomak ez baldin badituzte, sistemak ez du erantzun bat bueltatuko, ezta ausaz ere. Sistemak, ausaz erantzun bat bueltatuko du – egokiak izan daitezkeen artean – sintoma bat edo gehiago komunean baldin baditu.

Komunean dituzten sintomak zenbatzeko bi aukera desberdin aztertu dira: (1) bakarrik zenbatzea berdinak diren sintomak, (2) berdinak diren sintomak zenbatzeaz gain, baita Levenshtein 2 distantziara dauden sintomak zenbatzea. Erabaki da bigarrena erabiltzea horrela, "migraña" eta "migrañas" berdintzat hartuko lituzkeelako. Azaldutako prozesua guztia 4.7. irudian ikus daiteke.



4.7 Irudia: Aukera egokia iragartzeko sistema.

Iragartzeko prozesua zehaztuta dagoela egin diren esperimenduak azalduko dira. Kasu honetan gure ereduak edozein erantzun posible kopuru dituen galderekin lan egin dezake bere portaera guztientzat berdina delarik. Hori dela eta, esperimenduak gure datu guztien, CasiMédicos-MIR *dataset*-a, gainean egingo dira. Hala ere, 4.3.1.1. atalean azaldu den bezala, galdera bakoitzak dituen erantzun guztietan ez dira entitateak detektatu. Beraz, esperimenduan zehaztasuna lau modu desberdinetan kalkulatu da:

1. Erantzun zuzena iragarriko da datu guztien gainean, hau da, CasiMédicos-MIR *dataset*-ak dituen 508 kasuen gainean.
2. Erantzun zuzena iragarriko da bakarrik erantzun posible guztietan entitateak detektatu dituzten kasuetan. Hots, 5 erantzun posible dituzten kasuetan 5 erantzunetan entitateak detektatu dituzten galderetan eta 4 erantzun posible dituzten kasuetan 4 erantzunetan entitateak detektatu dituzten galderetan. Ingelesez egoera honetan dauden 73 kasu daude eta gaztelaniaz 198.
3. Erantzun zuzena iragarriko da 5 erantzun posible dituzten kasuetan 5 edo 4 erantzunetan entitateak detektatu dituzten galderetan eta 4 erantzun posible dituzten

kasuetan 4 edo 3 erantzunetan entitateak detektatu dituzten galderetan. Ingelesez egoera honetan dauden 122 kasu daude eta gaztelaniaz 252.

4. Erantzun zuzena iragarriko da erantzunen batean entitateak detektatu dituzten galderetan, hau da, 5 edo 4 erantzun posible izanda, gutxienez erantzun batean entitateak detektatu dituzten galderak. Ingelesez egoera honetan dauden 267 kasu daude eta gaztelaniaz 369.

### 4.4 MCQA hurbilpena

HP-ko atazak aurrera eramateko normalean hizkuntza-ereduak erabiltzen dira. Beraz, MIR azterketetako galderen erantzun egokia iragartzeko hizkuntza-eredu bat erabiltzen duen sistema bat eraikiko da, ondoren, ezagutza-baseekin eraikitako sistemarekin konparatzeko. Modu horretan, bakoitzak dituen abantailak eta desabantailak aztertuko dira.

Inplementazioari dagokionez [Pal et al.](#)-ren artikuluan atzigarri dagoen kodea erabiliko da erreferentzia moduan. Lehenik eta behin, sistemaren konfigurazioa eta artikuluan aurkezten den oinarritzko eredua azalduko da. Ondoren, egin diren esperimentuak eta horiek aurrera eramateko egin zaizkion aldaketak azalduko dira.

#### 4.4.1 Sistemaren konfigurazioa

Sistemaren konfigurazioari dagokionez *overfitting*-ari aurre egiteko *early stopping*<sup>11</sup> eta *dropout*<sup>12</sup> erregulazio teknika gehitu da. Bestalde, Titan V (12Gb) 2 GPU-tan hurrengo konfigurazioarekin eredu bakoitza entrenatuko da:

- Batch\_size = 4 ([Pal et al.](#)-ren lanean 32 zen)
- learning\_rate = 2e-4
- hidden\_dropout\_prob = 0.4
- epoch = 5
- early stopping-eko patience = 2

#### 4.4.2 Oinarritzko eredua

Erreferentziako artikuluan erabiltzen duten oinarritzko eredua erabiliko da. Oinarritzko ereduak MedMCQA *dataset*-a erabiltzen du eredua entrenatu eta ebaluatzeko. Ereduak BERT bat fintzen (*fine-tuned*) du MedMCQA-ko *train* eta *dev* partizioak erabiliz. Tamaina desberdinetako BERT-ak daude, geruza eta dimentsio kopuru desberdinak dituztelarik. Zenbat eta txikiago izan eredua, azkarrago entrenatzen da, baina lortzen diren emaitzak okerragoak izan ohi dira. Kasu honetan *bert-base-uncased* eredu handia erabiliko da 12 geruza, 768ko tamainako ezkutuko geruza eta milioi bat baino gehiagoko parametro dituena. Honetaz aparte, artikuluan medikuntzan espezializatuak dauden BERT desberdinak, hala

---

<sup>11</sup>[https://en.wikipedia.org/wiki/Early\\_stopping](https://en.wikipedia.org/wiki/Early_stopping)

<sup>12</sup>[https://en.wikipedia.org/wiki/Dilution\\_\(neural\\_networks\)](https://en.wikipedia.org/wiki/Dilution_(neural_networks))

nola SciBERT, BioBERT eta PubmedBERT erabiltzen dira ereduaren zehaztasuna hobetzeko, erakutsi baita medikuntza atazetan medikuntzan espezializatuak dauden BERT-ak erabiltzean emaitzak hobetzen dituela.

Erabili duten arkitektura 4.8. irudian ikus daiteke. Galdera bakoitzak daukan erantzun posible bakoitzeko, kasu honetan lau, BERT desberdinek erabiltzen duten Transformerrari sarrera moduan sekuentzia desberdin batzuk sortzen zaizkio. Sekuentzia horiek testuingurua erabili dezakete edo ez. Testuingurua erabiltzen ez baldin bada [CLS] Galdera [SEP] Aukera [SEP] sartzen zaio sekuentzia moduan erantzun bakoitzeko. [CLS] Testuingurua [SEP] Galdera [SEP] Aukera [SEP] sekuentzia aldiz, sartzen zaio erantzun bakoitzeko testuingurua erabiltzen bada.

Sekuentzian dagoen testuingurua *dataset*-ean agertzen den azalpena da, azalpen hori Wikipedia eta PubMed erabiliz sortu da *dataset*-a egin denean. Galdera bakoitzean dagoen erantzun bakoitzeko sortzen den sekuentzian erabiltzen den testuingurua berdina da guztietan, azalpen bakarra baitago galdera bakoitzean. Testuingurua erabiltzearen helburua ereduaren ikasketa prozesua laguntzea da honek uler dezan erantzun zuzenaren zergatia. Izan ere, ereduak aukera egokiena iragarri behar du erantzun posible bakoitzarekin emandako testuinguruaren arabera. Adibidez, 4.12. taulako kasua baldin badaukagu, testuingururik gabe erabiltzen bada, Transformerraren sarrera hurrengoa izango litzateke MedMCQA *dataset*-arekin:

- [CLS] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Cricoid cailage* [SEP]
- [CLS] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Thyroid cailage* [SEP]
- [CLS] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Cunieform cailage* [SEP]
- [CLS] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Epiglottis* [SEP]

Testuingurua erabiltzen bada, aldiz, hurrengoa izango litzateke. Aurrerago azalduko diren esperimuetan erabiliko den *CasiMédicos-MIR dataset*-arekin berdina izango litzateke testuingurua mediku bolondresen azalpenak izanik:

- [CLS] *Complete cailage ring is seen in Cricoid cailage. Cricoid cailage is shaped like a ring. It encircles the larynx below the thyroid cailage. It is thicker and stronger than the thyroid cailage.* [SEP] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Cricoid cailage* [SEP]
- [CLS] *Complete cailage ring is seen in Cricoid cailage. Cricoid cailage is shaped like a ring. It encircles the larynx below the thyroid cailage. It is thicker and stronger than the thyroid cailage.* [SEP] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Thyroid cailage* [SEP]
- [CLS] *Complete cailage ring is seen in Cricoid cailage. Cricoid cailage is shaped like a ring. It encircles the larynx below the thyroid cailage. It is thicker and stronger than the thyroid cailage.* [SEP] *Which of the following is the only complete cailage ring in the respiratory tree* [SEP] *Cunieform cailage* [SEP]
- [CLS] *Complete cailage ring is seen in Cricoid cailage. Cricoid cailage is shaped like a ring. It encircles the larynx below the thyroid cailage. It is thicker and stronger than the thyroid*

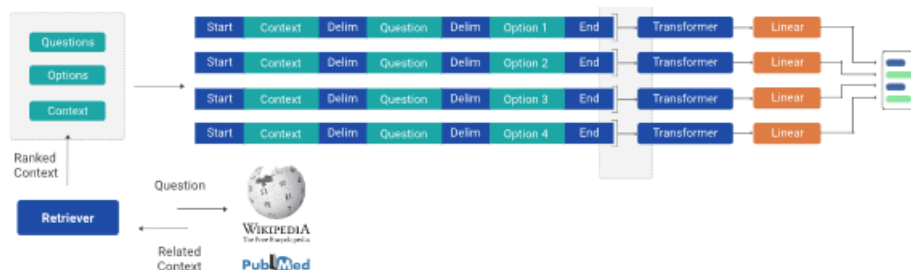
#### 4. GARATUTAKO INPLEMENTAZIOA

*cailage.* [SEP] *Which of the following is the only complete cailage ring in the respiratory tree*  
 [SEP] *Epiglottis* [SEP]

id	question	opa	opb	opc	opd	cop	choice_type	exp	subject_name	topic_name
b64a9cd7-d076-4c55-8be1-f9e44feca6cc	Which of the following is the only complete cailage ring in the respiratory tree?	Cricoid cailage	Thyroid cailage	Cunieforn cailage	Epiglottis	0	single	Complete cailage ring is seen in Cricoid cailage. Cricoid cailage is shaped like a ring. It encircles the larynx below the thyroid cailage. It is thicker and stronger than the thyroid cailage.	Anatomy	

4.12 Taula: MedMCQA dataset-aren adibide bat.

Ondoren, erantzun posible bakoitzeko sortutako sekuentzia Transformerretik pasatzen da. Amaitzeko, honek lortutako irteerako [CLS] tokenen gainean *softmax* geruza lineal bat aplikatu da aukera egokia iragartzeko, ataza hau sailkapen ataza bat moduan tratatu daiteke eta. Honek aukera bakoitzeko probabilitate bat bueltatzen du eta maximoa aukera egokia izango da.



4.8 Irudia: Oinarritzko ereduaren arkitektura [6].

#### 4.4.3 Egindako esperimentuak

Hasteko, Pal et al.-ren lanean aurkeztutako kodea ulertu ondoren bertan erabiltzen den *batch\_size* kopurua murriztu behar izatearen beharra ikusi da GPU-etan sartzen ez zelako. Hori dela eta, *batch\_size* 4ra murriztu da eta bere datuekin (bere *train* eta *dev* partizioekin) probatu da. Honen helburua, aldatutako ereduaren konfigurazioarekin lortzen diren emaitzak Pal et al.-ren lanean aurkeztutako emaitzekin asko aldatzen ez direla konprobatzea da. Horrela bada, esperimentuak egin eta lanean aurkeztutako emaitzekin konparatu ahal izango da eta bestela, konfigurazioa eta kodea aldatu behar izango dira antzeko emaitzak lortzeko.

Lanean 0,35ko *accuracy* lortzen da *dev* datuekiko (bere *test* partizioarekin ezin da probatu honen emaitzak ezkutututa daudelako) eta gure ereduarekin, *dev* partizio berdinearekin, 0,38ko *accuracy* lortu da. Emaitzak 0,03tan hobetu direnez, gure ereduarekin esperimentu desberdinak egin daitezke.

Egingo diren esperimentu guztiak *bert-base-uncased* BERT-arekin egingo dira nahiz eta oinarritzko artikuluan medikuntzan espezializatuak dauden BERT-ak ere erabili. Gure helburua medikuntzan espezializatu ez dagoen BERT batekin emaitzak hobetzea eta eredu gure beharretara moldatzea da. Gainera, BERT ez espezializatu batekin emaitzak hobetzen lortzen baldin badira, BERT espezializatuarekin ere hobetuko dira, honekin emaitza hobeak lortzen direla ondorioztatu baita.

Helburua MIR azterketetako erantzun zuzena iragartzea denez, CasiMédicos-MIR ingeleseko *dataset*-a erabiliko da. Hori dela eta, emango den lehenengo pausua *dataset*-a



moldatzea izango da 4.1.2. atalean agertzen den egitura izateko. Prozesu honetan aipagarri da erantzun zuzena adierazteko MedMCQA-n erantzunak 0 zenbakiarekin zenbakitzen hasten direla (0 eta 3 arteko zenbaki bat izanik), eta CasiMédicos-MIR corpusean 1 zenbakiarekin hasten direla zenbakitzen, horregatik -1 egin beharko da ondo funtziona dezan. Gainera, *train*, *dev* eta *test* partizioak egingo dira bakoitzak kasu guztien %60, %20 eta %20 izango dituztenak, hurrenez hurren. Baina CasiMédicos-MIR corpusean bost eta lau erantzun posible dituzten galderak daudenez, partizio guztietan lau eta bost erantzun posible, nahasturik, dituzten galderak egon behar dira ereduak ondo ikas dezan. Beraz, partizioak ausaz egingo dira *sklearn* liburutegiko *train\_test\_split* funtzioarekin.

Esan bezala, CasiMédicos-MIR corpusean bost erantzun posible dituzten galderak daude eta MedMCQA *dataset*-ean, aldiz, lau erantzun posible dituzten galderak bakarrik daude. Honek arazo bat suposatzen du MedMCQA ereduak ezin duelako galdera guztien erantzun zuzena iragarri. Hori dela eta, esperimendu desberdinak planteatuko dira alde batetik, agerian uzteko Pal et al.-ren lanak dituen mugak eta beste alde batetik, muga horiek gainditzeko eredu onena lortzeko. Jarraian egindako esperimendu desberdinak aurkeztuko dira. Hasiera batean azaltzen diren esperimenduetan ez da testuingurua erabiliko.

#### 4.4.3.1 Oinarrizko eredu (bert-base-uncased + MedMCQA + context = False)

Ikusteko Pal et al.-ren lanaren mugak zeintzuk izan daitezkeen, hasteko, oinarrizko ereduarekin probatuko da. Ereduak BERT-a fintzeko (*fine-tuned*) MedMCQA *dataset*-a erabiltzen du. Hala ere, eredu bakarrik lau erantzun posible dituzten galderekin erabilgarria da, MedMCQA *dataset*-ak lau erantzun posible dituzten galderez osatuta baitago. Hori dela eta, gure CasiMédicos-MIR corpuseko galdera guztien erantzun zuzena zein den iragartzeko ez da gai.

Oinarrizko ereduarekin hiru esperimendu desberdin egingo dira. MedMCQA *dataset*-eko *train* eta *dev* partizioak erabiliko dira entrenatzeko eta ebaluatzeko CasiMédicos-MIR *dataset*-a erabiliko da galderak lau erantzun posible bakarrik izanda. CasiMédicos-MIR corpusean modu desberdinetan probatuko da:

1. CasiMédicos-MIR *dataset* osoko bakarrik 4 erantzun posible dituzten galdera guztiak *test* moduan probatuko dira.
2. CasiMédicos-MIR corpusearekin egindako *test* partizioarekin probatuko da:
  - a) Corpusearekin egindako *test* partizioak 4 erantzun bakarrik dituzten kasuekin. Lehenik, corpusearekin egindako *test* partizioak soilik 4 erantzun dituen kasuekin ebaluatuko da.
  - b) Corpusearekin egindako *test* partizioak 4 erantzun bakarrik dituen kasuekin + 5 erantzun dituzten kasuetan erantzun bat ausaz kendu egokia ez bada.

Test partizioak totalen jada kasu gutxi (4.13 taulako 101 kasu) baldin baditu, 4 erantzun bakarrik dituzten kasu kopurua are gutxiago dira (4.13 taulako 49 kasu). Hori dela eta, bigarrenik, kasuak handitzeko, 5 erantzun dituzten kasuak moldatuko dira 4 erantzun bakarrik izateko. Zehazki, 5 erantzun dituzten kasu horietan erantzun bat ausaz kenduko da. Kendutako erantzuna egokia baldin bada, ez da beste bat ausaz aukeratu alborapenik ez sartzeko, bestela ez zen

#### 4. GARATUTAKO INPLEMENTAZIOA

izango guztiz ausazko prozesu bat. Kasu horiek ez dira ebaluatuko eta *accuracy* kalkulatzeko ez dira ontzat hartuko. Horrela, 96 kasu (ikus 4.13. taula) ebaluatuko dira eta *accuracy* kalkulatu da 101 kasuen gainean.

	4 eta 5 erantzun dituzten kasuak	4 erantzun bakarrik dituzten kasuak	4 erantzun + 5 erantzunetakoak 4koak bihurtu
<b>test</b>	101 kasu	49 kasu	96 kasu

4.13 Taula: CasiMédicos-MIR corpuseko *test* partizioan dauden kasu kopurua egoera desberdinetan.

#### 4.4.3.2 Oinarrizko ereduaren bert-a CasiMédicos-MIR corpusarekin findu (bert-base-uncased + CasiMédicos-MIR + context = False)

Orain arte egindako esperimentuekin, CasiMédicos-MIR corpuseko datuak eredu ebaluatzeko baino ez dira erabili. Beraz, gure datuek (CasiMédicos-MIR) entrenamenduan erabiltzean ereduaren zehaztasuna hobetzen laguntzen duten ikusi nahi da. Horretarako, gure datuekin egin diren *train* eta *dev* partizioak erabiliko dira entrenamenduan eta horiekin BERT-a findu egingo da. Partizio bakoitzean dauden kasuak gutxi direnez, jarraian azalduko den esperimentu bakoitza bost *seed* desberdinekin (42, 60, 50, 46, 55) probatuko da. Horrela, horien arteko desbiderapenak zenbatekoak diren ikus daiteke eta ereduaren zehaztasuna gutxi gorabehera zenbatekoa den finkatu ahalko da.

Aurreko esperimentuekin lortutako emaitzekin konparatu ahal izateko, alde batetik, egingo dira esperimentu batzuk 4 erantzun posible dituzten galderekin. Beste alde batetik, gure datuak erabiltzen direnez eta horietan partizio guztietan 4 eta 5 erantzun posible dituzten kasuak daudenez, beste esperimentu batzuk egingo dira 4 eta 5 erantzun posible dituzten galderekin.

Hasteko, **4 erantzun posible dituzten galderekin** egindako esperimentuak azalduko dira. Egingo diren esperimentuak bi *test* partizioa desberdinekin probatuko dira. (1) Esperimentuak egingo dira alde batetik, *test* partizioak 4 erantzun bakarrik dituen galderekin eta (2) bestetik, kasuak handitzeko 5 erantzun dituzten kasuetan erantzun bat ausaz kenduz. Azkenengo esperimentu horietan, aurrekoan bezala, 96 kasu ebaluatu eta 101 kasuen gainean *accuracy* kalkulatu da.

*Train* eta *dev* partizioei dagokienez, horiek ere 4 erantzun bakarrik izan behar dituzte. Hori dela eta, lehenik eta behin, *train* eta *dev* partizioetan 4 erantzun dituzten kasuak harturik eredu finduko da eta aipatutako *test* partizioa bi egoeratan probatuko da. *Train* eta *dev* partizioek 4 erantzuneko galdera gutxi dituztenez (4.14. taulako 162 eta 48 kasu, hurrenez hurren), kasu kopurua handitzeko 5 erantzun posible dituzten galderetan 5 erantzuna kentzea, betiere zuzena ez bada, erabaki da. Horrela, 4.14. taulan agertzen diren 274 eta 87 kasuetara handitu da eta partizio horiekin, berriz ere, eredu finduko da eta *test* partizioa bi egoeratan probatuko da.

Esperimentu horiekin, kasu batzuetan ikusi da *early stopping*-en finkatzen den *patience* balio baxuagatik, 2 zehazki, eredu gelditzen dela *epoch* kopurura ailegatu baino lehen. Nahiz eta *early stopping*-aren funtzioa hori izan, partizioetan dauden datu kopurua txikia denez, gelditzen denean ereduari oraindik ikasteko falta zaiolaren ustea dago. Beraz, esperimentuak errepikatuko dira gehiago entrenatzeko eta ondorioz, gehiago ikasteko.

Izan ere, esperimentuak *patience* 10 izanik eta *epoch* kopurua 20 izanik berriz egingo dira. Laburbilduz, hauek dira egingo diren esperimentuak:

1. Corpusarekin egindako *test* partizioak 4 erantzun bakarrik dituen kasuekin probatu da:
  - a) *train* eta *dev* partizioetan 4 erantzun dituzten kasuak harturik:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.
  - b) *train* eta *dev* partizioetan 4 erantzun dituzten kasuak harturik + 5. erantzuna kendu aukera egokia ez denean:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.
2. Corpusarekin egindako *test* partizioak 4 erantzun bakarrik dituen kasuekin + 5 erantzun dituzten kasuetan erantzun bat ausaz kendu egokia ez bada.
  - a) *train* eta *dev* partizioetan 4 erantzun dituzten kasuak harturik:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.
  - b) *train* eta *dev* partizioetan 4 erantzun dituzten kasuak harturik + 5. erantzuna kendu aukera egokia ez denean:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.

	4 eta 5 erantzun dituzten kasuak	4 erantzun bakarrik dituzten kasuak	4 erantzun + 5. erantzuna kendu aukera egokia ez denean
<b>train</b>	306 kasu	162 kasu	274 kasu
<b>dev</b>	101 kasu	48 kasu	87 kasu

**4.14 Taula:** CasiMédicos-MIR *train* eta *dev* partizioetan dauden kasu kopurua egoera desberdinetan.

Orain, **4 eta 5 erantzun posible dituzten galderekin** egindako esperimentuak azalduko dira. Hortaz, CasiMédicos-MIR corpusarekin egindako partizioak aldatu gabe ereduak findu eta ebaluatuko da.

Hori guztia egin ahal izateko kodea aldatu beharko da. Izan ere, Dataset klasean, hedagarria izateko asmoz, bi klase desberdin sortuko dira: MCQADataset5 eta MCQADataset4. Bien arteko diferentzia bakarra *options* atributua, *csv*-an dauden erantzun posibleen zutabeen izenak (opa, opb, opc, opd eta MCQADataset5 kasuan ope) da. Etorbizunean beste erantzun kopuru posible dituzten *dataset*-ekin lan eginez gero, beste klase bat baino ez da sortu behar *options* atributua aldatuz. Azkenik, `args.py` fitxategiko `num_choices` aldatuko da aukera posible kopurura (4 edo 5, gure kasuan).

CasiMédicos-MIR corpusak 4 erantzun posible dituzten kasuetan 5. erantzuna hutsik agertzen denez, kodean horrekin zer egin islatu behar da. Kasu honetan, 5. erantzuneko

sekuentzia ere Transformerraren sarrera moduan sartuko litzateke. Esperimentuei dagokie-  
nez probatuko da alde batetik, hutsik dauden erantzunak sekuentzian hutsik mantenduz eta  
beste alde batetik, hutsik dauden erantzunetan, dauden erantzun posibleen artetik luzeena  
hartu eta honen erantzunaren testuaren erdia jarritz. Adibidez, erantzun luzeena “Derivar a  
fisioterapia para elasticación del raquis.” baldin bada, 5. erantzun horretako sekuentziako  
erantzunean jarriko litzateke “Derivar a fisioterapia para ”. Horrela, galdera bakoitzeko  
hutsik dagoen aukera ez da beti berdina izango eta ereduak ez du ikasiko hutsik dauden  
erantzunak beti desegokiak direla. Gainera, 5. erantzun posiblea ez da errepikatua izango  
testuaren erdia jartzen delako.

Esperimentu horiek kasu honetan ere sistemaren konfigurazioa normalarekin eta hura  
*patience* 10era eta *epoch* kopurua 20ra aldatuz probatuko dira. Laburbilduz, hauek dira egin  
diren esperimentuak bost *seed* desberdinekin:

1. CasiMédicos-MIR corpusarekin egindako *train*, *dev* eta *test* partizioak 4 eta 5 erantzun  
dituztelarik probatu da:
  - a) Hutsik dagoen 5. erantzun posiblea hutsik mantenduz:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.
  - b) Hutsik dagoen 5. erantzun posiblea hutsik **ez** mantendu, hartu erantzun posi-  
bleen artetik luzeenaren testuaren erdia:
    - i. Sistemaren konfigurazioa normalarekin.
    - ii. Sistemaren konfigurazioa *patience* 10era eta *epoch* kopurua 20ra aldatuz.

#### 4.4.3.3 Oinarrizko eredia MedMCQA *dataset*-arekin eta CasiMédicos-MIR (**bert-base-uncased + MedMCQA + CasiMédicos-MIR + context = False**)

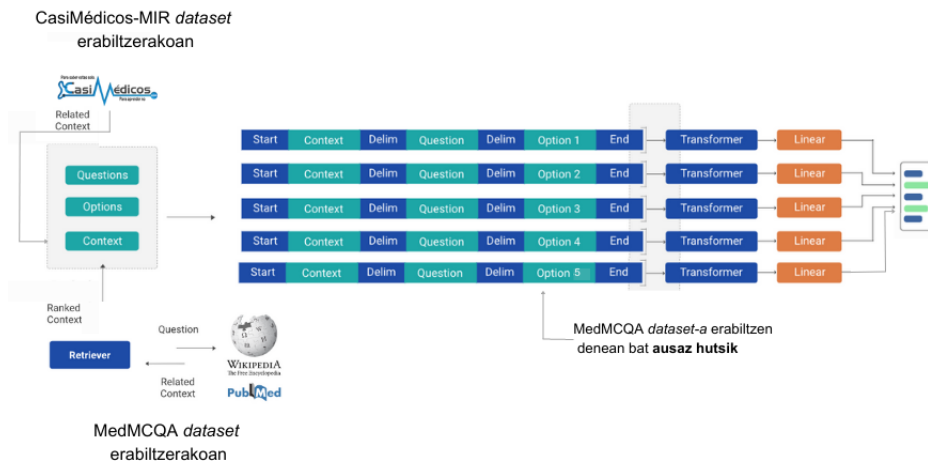
Nahiz eta aurreko ataleko esperimentuekin CasiMédicos-MIR datuak entrenamenduan  
erabili, ikusi nahi da ea datu kopuru handi baten gainean, MedMCQA gainean, entrena-  
tuta dagoen eredu handi bat CasiMédicos-MIR datuekin entrenatuz, zelako eragina duen  
ereduaren zehaztasunean.

Horretarako, aurreko atalean azaldu diren esperimentu berdinak egingo dira baina  
orain, 4.4.3.1. ataleko eredia CasiMédicos-MIR datuekin finduko da. Hori egiteko, kodeari  
dagokionez 4.4.3.1. atalean egindako esperimentuetatik *checkpoint* onena kargatu eta  
CasiMédicos-MIR datuak erabiliz gehiago jarriko da entrenatzen.

#### 4.4.3.4 Bost erantzuneko eredu berria + context = False

Aurreko atalean CasiMédicos-MIR corpusa erabili da oinarrizko eredia hobetzen saiatzeko.  
Baina hala ere, ereduak bost erantzun posible dituzten galderekin ez du balio. Hori dela  
eta, eredu berri bat sortuko da zeinak bost erantzun posible dituzten galderetan erantzun  
zuzena iragartzeko gai den.

Horretarako, hasteko, MedMCQA *dataset*-a aldatuko da 5 erantzun posible izateko,  
hutsik dagoen 5. erantzun bat ausazko posizio batean gehituko da. Horrela, 5. erantzuna  
beti ez da hutsik egongo eta ereduak ez du ikasiko hura beti desegokia dela. Erantzun  
huts bat gehitzerakoan kontuan izango da erantzun zuzenaren zenbakia aldatu behar dela



**4.9 Irudia:** Eredu berriaren arkitektura, 5 erantzun posible dituzten galderak erantzuteko gai dena.

kasu batzuetan. Adibidez, erantzun zuzena 2 bada eta sartzen bada 5. erantzun hutsa 1. posizioan, erantzun zuzena orain 3 izango da. Bestalde, MCQADataset5 motatako Dataset objektuak sortuko dira partizio bakoitzeko eta `args.py` fitxategiko `num_choices` aldatuko da 5 jarritz.

Gauzak horrela, aldatuta dauden MedMCQA *train* eta *dev* partizioak erabiliko dira BERT-a fintzeko eta *test* partizio moduan hurrengoa erabiliko da: 4 eta 5 erantzun dituen CasiMédicos-MIR corpus osoa eta 4 eta 5 erantzun dituen CasiMédicos-MIR corpusarekin sortutako *test* partizioa. Bi esperimentu horiek ere probatuko dira alde batetik, hutsik dauden erantzunak hutsik mantenduz eta beste alde batetik, hutsik dauden erantzunetan, dauden erantzun posibleen artetik luzeena hartu eta honen erantzunaren testuaren erdia jarriz. 4.9. irudian ikus daiteke eraikiko den eredu berria, kasu honetan *softmax* geruza lineala lortutako bost irteerako `[CLS]` tokenen gainean aplikatuko da. Arkitektura aurrekoaren antzekoa da baina 5. erantzuneko sekuentzia ere sarrera moduan sartuz. Laburbilduz hauek dira egin diren esperimentuak:

- Aldatuta dauden MedMCQA *train* eta *dev* partizioak eta *test* moduan CasiMédicos-MIR probatu da:
  1. Hutsik dagoen 5. erantzun posiblea hutsik mantenduz:
  2. Hutsik dagoen 5. erantzun posiblea hutsik **ez** mantendu, hartu erantzun posibleen artetik luzeenaren testuaren erdia.
- Aldatuta dauden MedMCQA *train* eta *dev* partizioak eta corpusarekin egindako *test* partizioarekin probatu da:
  1. Hutsik dagoen 5. erantzun posiblea hutsik mantenduz:
  2. Hutsik dagoen 5. erantzun posiblea hutsik **ez** mantendu, hartu erantzun posibleen artetik luzeenaren testuaren erdia.

#### 4.4.3.5 Testuingurua erabili (context = True)

Esperimentu guztiak testuingurua erabili gabe egin direnez, Pal et al.-ren artikuluan egin den bezala, ikusi nahi izan da testuinguruak nola laguntzen dion ereduari. Hori dela eta, esperimentu guztiak errepikatuko dira baina orain testuingurua erabiliz.

MedMCQA *dataset*-a erabili denean, Wikipedia eta PubMed erabili da testuingurua sortzeko. Baina CasiMédicos-MIR erabili denean, testuingurua CasiMédicos webgunean mediku bolondresek idatzi dituzten azalpenak dira, 4.9. irudian ikus daitekeenez. Kasu honetan galdera guztiek ere testuinguru berdina erabiliko dute. Adibidez, MedMCQA-k duen egitura berdina izan dezan, aldatutako CasiMédicos-MIR-eko corpuseko 4.15. taulako kasua hartzen badugu, testuingurua erabilia Transformerrak sarrera moduan jasoko litzkeen sekuentziak hurrengoak izango lirateke:

- [CLS] *The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.* [SEP] *An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:* [SEP] *Staphylococcus aureus.* [SEP]
- [CLS] *The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.* [SEP] *An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:* [SEP] *Streptococcus pneumoniae.* [SEP]
- [CLS] *The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.* [SEP] *An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:* [SEP] *Candida albicans.* [SEP]
- [CLS] *The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.* [SEP] *An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:* [SEP] *Pseudomonas aeruginosa.* [SEP]
- [CLS] *The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.* [SEP] *An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:* [SEP] *Aspergillus fumigatus.* [SEP]

id	question	opa	opb	opc	opd	ope	cop	choice_type	exp	subject_name	topic_name
2011-117	An immunodeficient patient who presents a pneumonia with meniscus halo sign or crescentic contour on chest X-ray/CT suggests infection by a microorganism:	Staphylococcus aureus.	Streptococcus pneumoniae.	Candida albicans	Pseudomonas aeruginosa.	Aspergillus fumigatus.	4	single	The halo sign is characteristic of pulmonary aspergillosis, especially in an immunodepressed patient. But it is not pathognomonic; it has also been associated with TB, some neoplasms and Wegener's granulomatosis.	Medicine	

**4.15 Taula:** MedMCQA-k duen egitura berdina izan dezan, aldatutako CasiMédicos-MIR-eko corpuseko adibide bat.

## Emaitzak eta eztabaida

Lanean bi hurbilpen desberdin, ezagutza-baseak eta hizkuntza ereduak, egin direnez eztabaida zatitu da. Egindako hurbilpen bakoitzeko lortutako emaitzak eta eztabaida aurkeztuko dira. Ondoren, bi hurbilpenak konparatuko dira horien emaitzak eta eztabaida aurkeztuz. Horrela, bi hurbilpenen abantailak eta desabantailak aztertuko dira.

### 5.1 Ezagutza-baseetan oinarritutako hurbilpena

Atal honetan hurbilpen honetan egindako esperimenduak azaltzen dira zehaztasuna era desberdinetan kalkulatzeko delarik. Emaitzei bider 100 egin zaie ulergarriagoak izan daitezen, 5.1. taulan emaitzak ikus daitezke, bertan parentesi artean agertzen da galdera guztien artetik asmatutako galdera kopurua.

Taulan agertzen diren emaitzak ulertzeko, aztertuko dugu sistemaren portaera asmatutako galderekin eta asmatu ez dituen galderekin. Asmatutako bi galderak 5.1. irudian ikus daitezke erantzun egokia asmatzeko gerta daitezkeen bi egoerekin. 5.1a. irudian aukera egokia asmatu du galderako sintomekin komunean gehien dituen erantzuna aurkitu duelako. 5.1b. irudian, aldiz, galderako sintomekin komunean sintoma kopuru berdina dituzten bi erantzun posible daude eta ausaz aukera egokia asmatu du. Hori dela eta, 5.1. taulan agertzen diren emaitzak *seed*-aren arabera aldatzen dira.

Asmatu ez diren galderak analizatuz gero, ez asmatzearen arrazoiak anitzak dira. Has-teko, hainbat galderen erantzun guztietan entitatearik ez da detektatu eta galdera horietan ezin izan da erantzun zuzena bueltatu. Beste batzuetan, entitateak detektatu dira baina ez erantzun posible guztietan. Beraz, sistemak galdera horietan erantzun gutxiren sintomak ditu galderetakoekin konparatzeko. Izan ere, egindako esperimenduetan asmatu diren galdera guztiak erantzun posible guztietan (5/5 edo 4/4), edo erantzun posible guztietako batean izan ezik (4/5 edo 3/4), entitateak detektatuta dituzten galderak izan dira.

Jarraitzeko, nahiz eta galdera askok erantzun posible guztietan entitateak detektatuta izan, askok eraikitako ezagutza-basean ez dituzte sintomarik edo ez daude ezagutza-basean. Gainera, entitateak detektatzen dituzten sistemen F1 balioa ez da %100koa. Hori dela eta, baliteke sintoma edo gaixotasun bat ez detektatzea edota detektatutakoa okerra izatea. Bestalde, galderetan sintoma gutxi daude eta horrek zailtzen du erantzunen sintomak

## 5. EMAITZAK ETA EZTABAIDA

		Ingelese	Ingelese NCI sintomekin	Gaztelaniaz	Gaztelaniaz erlazioekin
<b>Datu osoak</b> (508 kasu)	<b>seed 42</b>	1,77 (9)	1,38 (7)	1,18 (6)	1,57 (8)
	<b>seed 60</b>	2,17 (11)	1,77 (9)	1,57 (8)	1,77 (9)
	<b>seed 50</b>	1,97 (10)	1,97 (10)	1,57 (8)	1,77 (9)
	<b>seed 46</b>	1,97 (10)	1,77 (9)	1,38 (7)	1,97 (10)
	<b>seed 55</b>	1,57 (8)	1,18 (6)	1,18 (6)	1,38 (7)
<b>5/5 eta 4/4</b> (ingelesez 73 eta gaztelaniaz 198 kasu)	<b>seed 42</b>	6,85 (5)	5,48 (4)	3,03 (6)	3,54 (7)
	<b>seed 60</b>	8,22 (6)	6,85 (5)	4,04 (8)	4,04 (8)
	<b>seed 50</b>	6,85 (5)	8,22 (6)	4,04 (8)	4,04 (8)
	<b>seed 46</b>	8,22 (6)	8,22 (6)	3,54 (7)	4,55 (9)
	<b>seed 55</b>	5,48 (4)	4,11 (3)	3,03 (6)	3,03 (6)
<b>[4-5]/5 eta [3-4]/4</b> (ingelesez 122 eta gaztelaniaz 252 kasu)	<b>seed 42</b>	6,56 (8)	4,91 (6)	2,38 (6)	3,17 (8)
	<b>seed 60</b>	8,20 (10)	6,56 (8)	3,17 (8)	3,57 (9)
	<b>seed 50</b>	7,38 (9)	7,38 (9)	3,17 (8)	3,57 (9)
	<b>seed 46</b>	7,38 (9)	6,56 (8)	2,28 (7)	3,97 (10)
	<b>seed 55</b>	5,74 (7)	4,10 (5)	2,38 (6)	2,78 (7)
<b>[1-5]/5 eta [1-4]/4</b> (ingelesez 267 eta gaztelaniaz 369 kasu)	<b>seed 42</b>	3,37 (9)	2,62 (7)	1,63 (6)	2,17 (8)
	<b>seed 60</b>	4,12 (11)	3,37 (9)	2,17 (8)	2,44 (9)
	<b>seed 50</b>	3,75 (10)	3,75 (10)	2,17 (8)	2,44 (9)
	<b>seed 46</b>	3,75 (10)	3,37 (9)	1,90 (7)	2,71 (10)
	<b>seed 55</b>	3,00 (8)	2,24 (6)	1,63 (6)	1,90 (7)

**5.1 Taula:** Ezagutza-baseetan oinarritutako hurbilpenen galdera **guztien** artean lortutako zehaztasuna \*100 eginez hizkuntza desberdinetan. Zehaztasuna era desberdinetan kalkulatu da eta parentesi artean asmatutako erantzun kopurua agertzen da.

galderen sintometan aurkitzea. Zehazki, ingelesez 0,58 sintoma daude galdera bakoitzeko, hau da, galdera guztietan ez dira sintomak detektatu. Gaztelaniaz, aldiz, 2,9 sintoma daude galdera bakoitzeko

Azkenik, askotan gertatzen dena da erantzun posibleetan aurkitutako sintomak berdinak direla. Hots, erantzunetan agertzen diren gaixotasunek beraien artean oso sintoma antzekoak dituztela. Beraz, askotan erantzun aukera bakoitzeko daukan informazioa berdina edo ia berdina da eta ausaz asmatu behar du (galderarekin komunean sintomak baldin badituzte). Gainera, Levenshtein distantziaren atalasea 2 baino ez denez, erantzun eta galderako sintoma batzuk ez dira berdintzat hartzen. Adibidez, nahiz eta 5.1b. irudian agertzen den adibideko erantzuna asmatu, bertan ikus daitezke erantzun batek “fiebre periódica” eta besteak “fiebre continua” (irudian azpimarratuta) sintoma dutela eta galderan “fiebre” sintoma dagoela. Baina sistemak berdinak kontsideratzeko Levenshtein distantzia bakarrik neurtzen duenez, ez ditu berdintzat kontsideratzen.

Gauzak horrela, esan daiteke lortutako emaitzak ez direla guztiz esanguratsuak. Nahiz eta zehaztasuna modu desberdinetan kalkulatu, benetan sistemaren portaera ez dute islatzen. Sistemak galdera gutxi batzuetan baino ez du erantzun bat bueltatzen, hau da, gutxitan galdera eta erantzunetik sintomaren bat komunean daukate. Zehazki, hizkuntza bakoitzeko egindako esperimentuetan emaitza bat bueltatutako galdera kopurua 5.2. taulan agertzen da.



## 5.1. Ezagutza-baseetan oinarritutako hurbilpena

A 58-year-old hypertensive man comes to the emergency department for a 30-second **syncope** while jogging, with spontaneous recovery and no sequelae. Blood pressure is 135/65 mmHg. Cardiac auscultation shows a loud, intense systolic murmur, which is reduced with Valsalva maneuvers and an attenuated second tone. The ECG shows sinus rhythm at 72 bpm, with criteria of left ventricular hypertrophy and inverted T waves in anterior face. Point out the correct statement:

1. The clinical picture suggests **pulmonary thromboembolism**.
2. The data provided indicate **hypertrophic cardiomyopathy with severe obstruction** of the left ventricular outflow tract.
3. These data correspond to severe **aortic stenosis**. ✓
4. **Aortic dissection** should be excluded by contrast-enhanced computed tomography.

### SINTOMAK:

1. -
2. dyspnea, edema, fatigue, heart failure, pain, tachycardia, vertigo
3. angina pectoris, **syncope**
4. -

Un hombre de 58 años, hipertenso, acude a urgencias por un **sincope** de 30 segundos mientras corre, con recuperación espontánea y sin secuelas. La tensión arterial es 135/65 mmHg. En la auscultación cardíaca presenta un soplo sistólico rudo, intenso, que se reduce con maniobras de Valsalva y un segundo tono atenuado. El ECG muestra un ritmo sinusal a 72 lpm, con criterios de hipertrofia ventricular izquierda y ondas T invertidas en cara anterior. Señale la afirmación correcta:

1. El cuadro clínico sugiere un **tromboembolismo** pulmonar.
2. Los datos aportados indican **miocardiopatía hipertrófica** con **obstrucción** grave del tracto de salida de ventrículo izquierdo.
3. Estos datos corresponden a **estenosis aórtica** grave. ✓
4. Debe excluirse una **disecación de aorta** mediante una tomografía computerizada con contraste.

### SINTOMAK:

1. -
2. disnea, dolor, edema, insuficiencia cardíaca, taquicardia, vértigo
3. angina de pecho, **sincope**, angina de pecho, **sincope**
4. -

### (a) Komunean sintoma gehien daukan erantzuna aurkitu duenez asmatu du.

A 32-year-old woman has traveled to Cuba from where she returned three days ago. The day after her return, she goes to her family doctor for high fever, intense arthromyalgia and **headache** that had started before the return trip and he prescribes her paracetamol. Three days later and with no improvement, he presented in the morning with a generalized pruritic maculo-papular rash, which was more intense on the lower limbs where it developed into **petechiae**, for which he went to the emergency department. There is no relevant data in the EC except platelets 75000/mm<sup>3</sup> (htco 36%, leukocytes 4100 79% neutrophils). What is the most likely diagnosis?

1. **Infectious mononucleosis**.
2. Allergic exanthema.
3. **P. falciparum malaria**.
4. Typhoid fever.
5. **Dengue fever**. ✓

### SINTOMAK:

1. enlargement of lymph nodes, fever, hepatomegaly, leukocytosis, lymphadenopathy, lymphocytosis, rash, splenomegaly, tonsillitis
2. -
3. anemia, chills, coma, hepatomegaly, intermittent fever, jaundice, periodic fever, splenomegaly
4. abdominal distention, bradycardia, constipation, continuous fever, delusion, diarrhea, fatigue, gastrointestinal bleeding, **headache**, hepatomegaly, insomnia, intestinal perforation, leukopenia, pallor, prostration, rash, splenomegaly
5. adynamia, anorexia, arthralgia, bleeding, bradycardia, fatigue, **headache**, lymphadenopathy, maculopapular rash, myalgia, nausea, vomiting

Una **mujer** de 32 años ha viajado a Cuba de donde ha llegado hace tres días. Al día siguiente de su regreso acude a su médico de cabecera por **fiebre** elevada, artromialgias intensas y **cefalea** que le había comenzado antes de iniciar el viaje de regreso y le prescribe paracetamol. Tres días más tarde y sin haber experimentado mejoría presenta por la mañana un exantema maculo-papuloso pruriginoso generalizado que es más intenso en miembros inferiores en donde evoluciona a la formación de petequias por lo que acude a Urgencias. En las EC no se aprecian datos relevantes excepto las plaquetas 75000/mm<sup>3</sup> (htco 36%, leucocitos 4100 79% neutrófilos). ¿Cuál es el diagnóstico más probable?

1. **Mononucleosis** infecciosa.
2. Exantema alérgico.
3. **Malaria** por **P. falciparum**.
4. **Fiebre** tifoidea.
5. **Dengue**. ✓

### SINTOMAK:

1. -
2. eflorescencia
3. Esplenomegalia, anemia, coma, escalofrío, **fiebre periódica**, hepatomegalia, ictericia
4. Esplenomegalia, bradicardia, cansancio, **cefalea**, delirio, diarrea, distensión abdominal, eflorescencia, estreñimiento, **fiebre continua**, hemorragia gastrointestinal, hepatomegalia, insomnio, leucopenia, palidez
5. adenopatía, adinamia, anorexia, artralgia, bradicardia, cansancio, **cefalea**, erupción maculopapular, hemorragia, mialgia, náusea, vómito

### (b) Komunean sintoma gehien dituen 2 erantzun aurkitu ditu eta ausaz zuzena asmatu du.

**5.1 Irudia:** Asmatutako bi galdera eta bere erantzun posibleak, urdinez galderetan aurkitu diren sintomak eta gorritz erantzunetan aurkitu diren gaixotasunak agertzen dira. Erantzun bakoitzeko gaixotasunen sintomak ere agertzen dira komunean dituzten sintomak beltzez daudelarik. ✓: erantzun egokia adierazten du.

Gainera, 5.1. irudian azaltzen den bezala, batzuetan erantzun zuzena, posibleen artean, ausaz bueltatu eta asmatzen du. Hori dela eta, kalkulatu da zenbat galderetan **ez** dituen ausaz emaitza bat bueltatu eta horiek 5.2. taulan ikus daitezke. Taula horretan ere ikus daitezke ausaz emaitza bat bueltatu **ez** dituen horien artetik zenbat asmatu dituen.

Guzti hori jakinda, 5.3. taulan zehaztasuna beste era batean dago kalkulatu. Taula horretan lehenengo errenkadan ausazko emaitza bat bueltatu **ez** dituen artetik, asmatutakoen zehaztasuna kalkulatu da (hori ahalik eta altuena izatea nahi dugu). Horretarako hurrengo kalkulua egin da:  $\frac{\text{Ausazko emaitza bueltatu EZ dituen asmatutakoak}}{\text{Ausazko emaitza bueltatu EZ dituen galderak}}$ . Hurrengo errenkadan, aldiz, emaitza bat bueltatuko erantzun guztien artetik ausaz asmatutakoak **ere** kontuan hartzen dira zehaztasuna kalkulatzeko, hortaz, *seed* desberdinekin egin da proba. Hau da egin den kalkulua:  $\frac{\text{Asmatutako galderak}}{\text{Emaitza bat bueltatutako galdera kopurua}}$ . Beltzez hizkuntza bakoitzeko emaitza onenak daude eta 5.1. taulako emaitzekin konparatu ahal izateko emaitzak 100ekin biderkatu dira. Azkenik, azkenengo errenkadan *seed* desberdinekin lortutako emaitzen batezbestekoa eta desbiderapen estandarra kalkulatu dira.

Emaitzak ikusirik, ingelesez ez dira espero ziren emaitzak lortu. Espero zen NCI bidez lortutako emaitzek, sintoma gehiago lortzearen ondorioz, NCI gabe lortutakoak hobetzea. Baina ez da hori gertatu NCI bidez lortutako sintomak erredundanteak direlako, hau da, batzuetan errepikatuta daude, 5.2. irudian bezala, edo oso antzekoak dira. Esan daiteke NCI

## 5. EMAITZAK ETA EZTABAIDA

	Ingelesez	Ingelesez NCI sintomekin	Gaztelaniaz	Gaztelaniaz erlazioekin
Emaizta bat bueltatuko galdera kopurua	31	31	30	35
Ausazko emaitza bueltatu EZ dituen galderak	21	21	23	28
Ausazko emaitza bueltatu EZ dituen asmatutakoak	7	6	6	7

5.2 Taula: Hizkuntza bakoitzeko bueltatutako emaitzen galderen inguruko informazioa.

	Ingelesez	Ingelesez NCI sintomekin	Gaztelaniaz	Gaztelaniaz erlazioekin	
Ausazko emaitza bueltatu ez dituenak	33,33	28,57	26,09	25,00	
Emaizta bat bueltatutako guztiak	seed 42	29,03 (9)	22,58 (7)	20,00 (6)	22,85 (8)
	seed 60	<b>35,48</b> (11)	29,03 (9)	<b>26,67</b> (8)	25,71 (9)
	seed 50	32,25 (10)	<b>32,25</b> (10)	<b>26,67</b> (8)	25,71 (9)
	seed 46	32,25 (10)	29,03 (9)	23,33 (7)	<b>28,57</b> (10)
	seed 55	25,8 (8)	19,35 (6)	20,00 (6)	20,00 (7)
<b>Batezbestekoa</b>	<b>30,96</b>	<b>26,45</b>	<b>23,33</b>	<b>24,57</b>	
<b>Desbiderapen estandarra</b>	<b>3,68</b>	<b>5,30</b>	<b>3,24</b>	<b>3,26</b>	

5.3 Taula: Sistemak bueltatutako erantzunak bakarrik kontuan harturik lortutako zehaztasuna \* 100 ginez hizkuntza desberdinetan. Parentesi artean asmatutako galdera kopurua agertzen da.

A 7-month-old infant comes to the emergency department referred by his pediatrician for suspected Kawasaki disease. He reports a **fever** of up to 39.5°C of 5 days evolution that does not subside despite treatment with amoxicillin prescribed 72 hours ago for suspected acute pharyngotonsillitis. During the last 24 hours he has had erythematous and pruritic **exanthema** of cephalocaudal evolution and during the last few days his parents have noticed a **cough** and intense conjunctivitis. No **rhinorrhea** was reported. Clinical examination showed subcrepitant in both lung bases without tachypnea or signs of **respiratory distress**. On oropharyngeal examination there are whitish spots in the jugal mucosa. What is the most probable diagnosis?

### SINTOMAK INGELESEZ:

1. diarrhea, **fever**, inflammation, lymphadenopathy, strawberry tongue
2. Pastia's lines, abdominal pain, fever, lymphadenopathy, rash, ring of paleness around the mouth, sore throat, strawberry tongue, tachycardia, tonsillitis, vomiting
3. -
4. enlargement of lymph nodes, **fever**, hepatomegaly, leukocytosis, lymphadenopathy, lymphocytosis, rash, splenomegaly, tonsillitis
5. Koplik's spots, Stimson's triad, anorexia, conjunctivitis, **cough**, diarrhea, **fever**, lymphadenopathy, maculopapular rash, runny nose ✓

### SINTOMAK INGELESEZ-NCI:

1. diarrhea, **fever**, inflammation, lymphadenopathy, strawberry tongue
2. Pastia's lines, abdominal pain, fever, lymphadenopathy, rash, ring of paleness around the mouth, sore throat, strawberry tongue, tachycardia, tonsillitis, vomiting
3. -
4. **Atypical Lymphocytes Present in Peripheral Blood, Diffuse Immunoblastic Proliferation, Fever, Lymphadenopathy, Non-Malignant Cellular Infiltrate, Polymorphic Cellular Infiltrate, Splenic Rupture, Splenomegaly**, enlargement of lymph nodes, **fever**, hepatomegaly, leukocytosis, lymphadenopathy, lymphocytosis, rash, splenomegaly, tonsillitis ✓
5. Koplik's spots, Stimson's triad, anorexia, conjunctivitis, **cough**, diarrhea, **fever**, lymphadenopathy, maculopapular rash, runny nose ✓

5.2 Irudia: Ingelesez eta ingelesez NCI bidez detektatu diren sintomak. Urdinez, galderan detektatutako sintomak eta berdez, NCI bidez lortutako sintoma berriak. Beltzez, komunean dituzten sintomak. NCI bidez, *fever* birritan agertzen denez, bi erantzunen artean berdinketa dago.

sintomek zarata baino ez dutela sartzen eta horren ondorioz, emaitzak okertzen direla.

Gaztelaniaz, aldiz, lortutako emaitzak bai dira esperotakoak hots, erlazioen bidez sintomen izen gehiago lortu direnez emaitzak hobetu dira. Alde batetik, emaitza bat bueltatutako galdera kopurua eta ausazko emaitza bat bueltatu **ez** dituen galdera kopurua handitu da (ikus 5.2. taula). Sintoma gehiago izateak, ingelesez ez bezala, informazio baliagarria ematen dio sistemari. Gainera, 5.2. taulan ikus daiteke emaitza bat bueltatu dituen galdera kopurua 5 unitate handitu dela. Zehazki, bost unitate horiek ausazko emaitza bueltatu **ez** diren galdera kopuru berriak dira, hori 5 unitatetan ere handitu delako. Hala ere, nahiz eta 5 unitate handitu, ausazko emaitza bat bueltatu **ez** dituen artean, asmatutakoen kopurua ez da hainbeste handitu, 1ean baino ez. Horrek esan nahi du, sistemak aukera bat bueltatzeko sintoma gehiago dituela baina, betiere sintoma horiekin ez dela aukera zuzena bueltatzen.

Bestalde, ingelesezko eta gaztelaniazko emaitzak konparatuz, ingelesez emaitza hobeak lortu dira. Nahiz eta ingelesez galdera bakoitzeko dauden sintomak gutxiago izan, inge-

lesezko ezagutza-basea aberatsagoa da eta MetaMap-ek detektatutako entitateak hobeak dira.

Azkenik, esan daiteke 5.3. taulan aurkeztutako emaitzak hobeak direla. Guztietan ausaz aukera bat aukeratuz gero, asmatzeko probabilitatea (5 galdera egonda %20) gaintzen dute. Hala ere, 5.1. eta 5.3. tauletako emaitzak ikusten badira agerian geratzen dira eraikitako ezagutza-basearen mugak. Nahiz eta eraikitako ezagutza-basea hizkuntza askotan eta ezagutza-base askotako informazioa osagarri izan, honen estaldura txikia da. Gaixotasun bakoitzeko sintoma gutxi lortzen dira eta askotan ez dira ezta lortzen ere.

## 5.2 MCQA hurbilpena

Hurbilpen honetan hainbat esperimendu egin dira. Bakoitzeko lortutako emaitzak jarraian aurkeztuko dira. Esperimendu guztiak errepikatu direnez testuingurua erabilia eta hura erabili gabe, emaitzak konparatzeko taula berdinean biak jarriko dira. Tauletan w/o jartzen duenean testuingurua ez dela erabili esan nahi du eta w jartzen duenean, aldiz, testuingurua erabili dela esan nahi du.

### 5.2.1 Oinarrizko ereduak (bert-base-uncased + MedMCQA)

Oinarrizko ereduarekin, hau da, BERT bat MedMCQA datuen gainean fintzean lortutako emaitzak 5.4. taulan ikus daitezke. Pal et al.-ren lanean bere *test* partizioarekin, testuingururik gabe lortutako emaitza 0,33koa izan da. CasiMédicos-MIR datuekin testuingururik gabe probatuz, antzeko emaitzak lortu dira 0,01eko diferentzia baino ez dagoelarik. Honek esan nahi du ereduak antzeko errendimendua daukala *test* partizioa aldatuz.

Testuingurua erabiltzen baldin bada, aldiz, lanean [6] lortutako 0,33 edo 0,37ko emaitzak ez dira lortu. Esperimendu batean emaitzak hobetzea lortu dira eta beste bi esperimenduetan, aldiz, ez dira hobetzea lortu. Gainera ikus daiteke testuingurua erabiltzeak emaitzak orokorrean okertzen dituela. Eredua entrenatzeko erabili den *dataset*-ean dagoen testuingurua PubMed eta Wikipediatik sortua izan da. Baina ereduak horren gainean ikasten duenak ez dio CasiMédicos-MIR datuen gainean iragarpen egoki bat egiteko balio.

<i>test</i> -ean erabilitako datu-multzoa		w/o test	w test
4 erantzun bakarrik dituzten guztiak		0,32	0,25
gure <i>test</i> -a	4 erantzun	0,32	0,44
	4 erantzun eta 5ekoetan ausaz bat kenduta	0,34	0,28

5.4 Taula: Oinarrizko ereduarekin (bert-base-uncased + MedMCQA) lortutako zehaztasuna.

### 5.2.2 Oinarrizko ereduaren bert-a CasiMédicos-MIR corpusarekin findu (bert-base-uncased + CasiMédicos-MIR)

Oinarrizko ereduak erabiltzen duen *bert-base-uncased* CasiMédicos-MIR corpusarekin fintzeko galdera batek izan ditzakeen erantzun kopuru kontuan izanik bi aukera aztertu dira: (1) galderek 4 aukera posible dituzte eta (2) galderek 4 eta 5 aukera posible dituzte.

### 5.2.2.1 4 aukera posible dituzten galderen esperimenduak

Oinarrizko ereduak erabiltzen duen *bert-base-uncased* CasiMédicos-MIR corpusarekin finduz lortutako emaitzak 5.5. taulan daude. Taula horretan agertzen diren esperimenduetan erabili diren galderek 4 aukera izan dituzte. Esperimendu desberdinak egin dira, lehenik, *train* eta *dev* partizioetan 4 erantzun dituzten kasuak hartu dira finketa egiteko. Ondoren, *train* eta *dev* partizioetan 4 erantzun dituzten kasuak hartuta eta 5 erantzun dituzten kasuak 4 erantzun posible izateko moldatuz, fintzeko erabili dira. Taulan kasu bakoitzeko konfigurazio normalarekin eta konfigurazioa aldatuta (*epoch*=20, *patience*=10) lortutako batezbestekoa eta desbiderazio estandarra agertzen dira. Bestalde, esperimenduak *test* partizio desberdinekin egin dira. “4” jarrita duten zutabeetan, *test* partizioak 4 erantzun bakarrik dituen kasuak ditu. “5to4” jarrita duten zutabeetan, aldiz, *test* partizioak 4 erantzun bakarrik dituen kasuak eta 5 erantzun dituzten kasuak 4 erantzun posible izateko moldatutako kasuak ditu. Erabilitako corpusa txikia denez, esperimendu guztiak 5 *seed* desberdinekin egin dira. *Seed* bakoitzarekin lortutako emaitzak A. eranskinean agertzen dira.

Train eta dev-eko 4ko erantzunak		w/o test		w test	
		4	5to4	4	5to4
Konfigurazio normala	Batezbestekoa	0,23	0,24	0,25	0,17
	Desbiderapen estandarra	0,09	0,06	0,09	0,02
Epoch=20, patience=10	Batezbestekoa	0,24	0,23	0,25	0,18
	Desbiderapen estandarra	0,06	0,04	0,07	0,02

Train eta dev-eko 4ko erantzunak eta 5koetan azkena kendu		w/o test		w test	
		4	5to4	4	5to4
Konfigurazio normala	Batezbestekoa	0,21	0,27	0,24	0,17
	Desbiderapen estandarra	0,12	0,05	0,08	0,06
Epoch=20, patience=10	Batezbestekoa	0,27	0,26	0,23	0,20
	Desbiderapen estandarra	0,12	0,06	0,08	0,04

**5.5 Taula:** Oinarrizko ereduak erabiltzen duen *bert-base-uncased* CasiMédicos-MIR 4 aukera dituzten galderekin finduz lortutako emaitzen batezbestekoa eta desbiderapen estandarra.

Lehenik testuingururik gabe lortutako emaitzak aztertuko ditugu. Ondoren, testuinguruarekin lortutakoak eta azkenik, biak konparatuko ditugu.

#### Testuingurua ez erabiltzearen eragina

- **“4” eta “5to4”-ren eragina.** Hasteko, “4” eta “5to4” zutabeetako batezbestekoak aztertzen baldin baditugu, konfigurazioaren arabera portaera desberdina dela ikus daiteke. Konfigurazio normalarekin egindako esperimenduetan “5to4” zutabeetako emaitzak hobeak dira. Baina konfigurazioa aldatzerakoan, aldiz, ez dira emaitzak hobetzen. Hala ere, 0,01eko batezbesteko diferentzia baino ez dago. “5to4” esperimenduetan *test* partizioa handiagoa da beraz, ereduak datuen distribuzioaren lagin handiago bat du. Lortutako emaitzek ereduaren errealitateko errendimenduaren adierazpen egokiagoa islatzen dute; izan ere, desbiderapen estandarrak konparatuz, baxuagoa da.
- **Konfigurazioaren eragina.** Bestetik, konfigurazio bakoitzarekin lortutako emaitzak konparatuz, ikus daiteke “4” zutabeetan emaitzak hobetzen direla *epoch* eta *patience* balioak handitzen direnean. Hots, gure ereduak datu gutxiren gainean entrenatu

denez, konfigurazio normalean definitu diren 5 *epoch*-ekin ez du behar beste ikasten, oraindik ikasteko falta zaio. Hori dela eta, gehiago entrenatzeak hobetzen ditu ereduaren emaitzak. Hala ere, “5to4” esperimentuetan non kasu gehiago dauden, emaitzak 0,01ean okertzen dira. Baina orokorrean gehiago entrenatzearen ondorioz, desbiderapen estandarra baxuagoa da.

- **Train eta dev partizioen kasu kopuruaren eragina.** Azkenik, *train* eta *dev* partizioak aldatuz lortutako emaitzak aztertuko ditugu. *Train* eta *dev*-eko 4 erantzunak dituzten galderak eta 5-koak 4 erantzun izateko moldatuz fintzeko erabiltzen badira, emaitza hobekien dira entrenatzeko datu gehiago daudelako. Horrela, ereduak datuen errepresentazio hobea eta orokortzeko gaitasun altuagoa dauka eta emaitza hobekien dira.

### Testuingurua erabiltzearen eragina

Testuinguruarekin lortutako emaitzak aztertuz gero, portaera ez da guztiz berdina. Alde batetik, esperimentu guztiekin “5to4” zutabetako emaitzak “4” zutabetako emaitzak baino txarragoak dira, testuingururik gabe ez bezala. Beste alde batetik, konfigurazioa aldatzeak, hau da, gehiago entrenatzeak ez ditu emaitzak asko hobetzen. Izan ere, emaitza gehienak berdinak edo 0,01ean hobetzen edo okertzen dira. Azkenik, *train* eta *dev* partizioak aldatuz lortutako emaitzak aztertuz, emaitza hobekien dira *train* eta *dev* partizioan kasu gutxiago daudenean (4ko galderak bakarrik). Ikus daitekeenez, testuingururik gabe lortutakoaren kontrakoa gertatzen da.

Testuinguruarekin lortutako portaera ez zen espero zena, [Pal et al.](#)-ren lanean testuinguruak emaitzak pixka bat hobetzen baitzituen. Honen arrazoia izan daiteke gure datuen, CasiMédicos-MIR datuen, testuinguruaren kalitatea MedMCQA datuen testuinguruarena baino txarragoa delako eta ondorioz, ereduak lagundu baino okertzen duelako. CasiMédicos-MIR datuen testuingurua mediku bolondresen idatzi dutenez honen egokitasuna zalantzarria da. Gainera, askotan testuinguruan esaten da zein den aukera egokia eta zeintzuk diren desegokiak baina, aukera horiek 1-5 artean daude eta ereduak aukerak 0-4 artean zenbakitzen ditu. Bestalde, *train* eta *dev* partizioak aldatuz lortutako emaitzak konparatuz, emaitza hobekien dira *train* eta *dev* partizioan kasu gutxiago daudenean (4ko galderak bakarrik) lortzea ere testuinguruagatik gertatzen da. *Train* eta *dev* partizioak aldatzen direnean 5 aukera dituzten galderak 4 izateko moldatuz, azkenengo aukera kentzen da. Beraz, eredu nahasi egiten da, aipatzen diren aukera desegokiaren artean, existitzen ez den 5. erantzuna aipatzen delako. MedMCQA-n, aldiz, testuingurua era egokiago batean sortu dute hainbat iturri medikotik ateraz informazioa.

### Testuingurua bai eta testuingurua ez erabiltzea

Testuinguruaren eragina aztertuz, ezin da ondorioztatu beti testuingurua erabiltzeak eredu hobetzen duenik. Batzuetan emaitzak hobetzen dira, baina beste batzuetan ez.

#### 5.2.2.2 4 eta 5 aukera posible dituzten galderen esperimentuak

Orain arte egindako esperimentuetan 4 aukera izan dituzte galderak, baina CasiMédicos-MIR 4 eta 5 aukera dituzenez, oinarritzko ereduaren erabiltzen den *bert-base-uncased* finduko da, baina *train* eta *dev* corpus osoekin. 5.6. taulan konfigurazio normalarekin eta gehiago

entrenatzeko konfigurazioa aldatuz egindako esperimenduekin lortutako emaitzak aurkezten dira. Bestalde, “5. hutsik” zutabean 4 aukera dituzten galderetan bosgarrena hutsik utziz lortutako emaitzak agertzen dira. “5. EZ hutsik” zutabean, aldiz, 4 aukera dituzten galderetan bosgarrena hutsik ez utziz lortutako emaitzak agertzen dira. Esperimendu hauek ere bost *seed* desberdinekin egin dira, emaitz horiek A. eranskinean aurki daitezke.

Train eta dev osoa		w/o test		w test	
		5. hutsik	5. EZ hutsik	5. hutsik	5. EZ hutsik
Konfigurazio normala	Batezbestekoa	0,25	0,27	0,18	0,19
	Desbiderapen estandarra	0,05	0,03	0,04	0,03
Epoch=20, patience=10	Batezbestekoa	0,26	0,22	0,21	0,26
	Desbiderapen estandarra	0,03	0,04	0,04	0,07

**5.6 Taula:** Oinarrizko ereduak erabiltzen duen *bert-base-uncased* CasiMédicos-MIR corpus osoarekin, 4 eta 5 aukera dituzten galderekin, finduz lortutako batezbestekoa eta desbiderapen estandarra.

- **Konfigurazioaren eragina.** Esperimendu hauetan, aurrekoan bezala, emaitzak orokorrean hobetzen dira eredu gehiago entrenatuz. Kasu honetan hobetzen dira bai testuinguruarekin bai testuingururik gabe. Nahiz eta esperimendu hauetan *train* eta *dev* partizioetan kasu gehiago egon, oraindik ere kasu gutxi dira eta ereduak gehiago entrenatzea behar du gehiago ikasteko.
- **Bosgarren erantzunaren eragina.** Bestalde, 5. erantzuna hutsik ez egoteagatik emaitza txarragoak lortuko direnaren hipotesia ez da bete. Izan ere, ia esperimendu guztietan emaitza hobeak lortu dira. Badirudi ereduak ez duela ikasten 5. aukera hutsik dagoenean aukera hori egokia ez dela. Edo nahiz eta ikasi, beste 4 aukeren artean egokia iragartzen ez dakiela. “5. EZ hutsik” zutabean 5. erantzuna hutsik dagoenean, beste aukeren arteko aukera luzeena hartu eta esaldi horren erdia jarri da. Hortaz, ere gerta daiteke kasu horretan ereduak ikastea bi aukeretan esaldi berdinak daudenean, horiek egokiak ez direla eta beraz, 3 aukeren artean iragarri behar izatea, horrela, asmatzearen probabilitatea igoz. Hala ere, esaldi luzeena aukera egokiarena izan daitekeenez, ereduak aukera egokia baztertzen ikas dezake.

Amaitzeko, aurreko esperimenduekin bezala, kasu gehienetan testuingurua erabiltzeak ez du eredu hobetzen. Honen arrazoia aurreko esperimenduen berdina izan daiteke, hau da, testuinguruaren kalitatea ona ez izatea.

### 5.2.3 Oinarrizko eredu MedMCQA dataset-arekin eta CasiMédicos-MIR (bert-base-uncased + MedMCQA + CasiMédicos-MIR)

MedMCQA datuekin finketa egin den oinarrizko modeloari CasiMédicos-MIR datuekin fintzeko galdera batek izan ditzakeen erantzun kopuru kontuan izanik, aurrekoan bezala, bi aukera aztertu dira: (1) galderak 4 aukera posible dituzte eta (2) galderak 4 eta 5 aukera posible dituzte.

#### 5.2.3.1 4 aukera posible dituzten galderen esperimenduak

MedMCQA datuekin findu egin den eredu CasiMédicos-MIR datuekin finduz zein izan den bere portaera 5.7. taulan ikus daiteke. Lehenengo errenkadan findu egingo den ereduaren

zehaztasuna dago adierazita, beste errenkada eta zutabeek 5.5. taulan adierazten dutenaren berdina esan nahi dute. Beltzez daude findu den ereduaren emaitzak hobetzea lortu duten esperimintuen zehaztasuna. Kasu honetan *seed* desberdinekin lortutako emaitzak, batez-bestekoak eta desbiderapenak daude. Hobekuntza hain txikia denez, ereduaren portaera ulertu ahal izateko.

		w/o test		w test	
		4	5to4	4	5to4
<b>bert-base-uncased + MedMCQA emaitzak</b>		<b>0,32</b>	<b>0,34</b>	<b>0,44</b>	<b>0,28</b>
Train eta dev-eko 4ko erantzunak, konfigurazio normala	seed 42	0,32	<b>0,35</b>	0,44	0,26
	seed 60	0,32	0,33	0,44	0,24
	seed 50	0,32	<b>0,35</b>	0,44	0,22
	seed 46	0,32	<b>0,35</b>	0,44	0,24
	seed 55	0,32	0,33	0,44	0,24
<b>Batezbestekoa</b>		<b>0,32</b>	<b>0,34</b>	<b>0,44</b>	<b>0,24</b>
<b>Desbiderapen estandarra</b>		<b>0,00</b>	<b>0,01</b>	<b>0,00</b>	<b>0,01</b>
Train eta dev-eko 4ko erantzunak, epoch=20, patience=10	seed 42	0,32	<b>0,35</b>	<b>0,48</b>	<b>0,33</b>
	seed 60	0,32	0,33	<b>0,56</b>	0,28
	seed 50	0,32	<b>0,35</b>	0,44	<b>0,33</b>
	seed 46	0,32	<b>0,35</b>	0,44	0,26
	seed 55	0,32	0,33	0,44	0,26
<b>Batezbestekoa</b>		<b>0,32</b>	<b>0,34</b>	<b>0,47</b>	<b>0,29</b>
<b>Desbiderapen estandarra</b>		<b>0,00</b>	<b>0,01</b>	<b>0,05</b>	<b>0,04</b>
Train eta dev-eko 4ko erantzunak eta 5koetan azkena kendu, konfigurazio normala	seed 42	0,32	0,31	0,44	0,26
	seed 60	0,32	<b>0,35</b>	0,44	0,24
	seed 50	0,32	<b>0,35</b>	0,44	0,24
	seed 46	<b>0,40</b>	0,33	0,44	0,24
	seed 55	0,32	<b>0,35</b>	<b>0,48</b>	0,24
<b>Batezbestekoa</b>		<b>0,34</b>	<b>0,34</b>	<b>0,45</b>	<b>0,24</b>
<b>Desbiderapen estandarra</b>		<b>0,04</b>	<b>0,02</b>	<b>0,02</b>	<b>0,01</b>
Train eta dev-eko 4ko erantzunak eta 5koetan azkena kendu, epoch=20, patience=10	seed 42	0,32	0,31	<b>0,48</b>	0,26
	seed 60	0,32	<b>0,35</b>	<b>0,48</b>	0,24
	seed 50	0,32	<b>0,35</b>	<b>0,48</b>	0,22
	seed 46	<b>0,40</b>	0,33	0,44	0,24
	seed 55	0,32	<b>0,35</b>	<b>0,48</b>	0,26
<b>Batezbestekoa</b>		<b>0,34</b>	<b>0,34</b>	<b>0,47</b>	<b>0,24</b>
<b>Desbiderapen estandarra</b>		<b>0,04</b>	<b>0,02</b>	<b>0,02</b>	<b>0,02</b>

5.7 Taula: Oinarritzko eredu MedMCQA dataset-arekin eta CasiMédicos-MIR corpusarekin finduz lortutako zehaztasuna. Beltzez emaitzak hobetzea lortu diren esperimintuak eta emaitzak.

### Testuingurua ez erabiltzearen eragina

Testuingururik gabe lortu diren emaitzak aztertuz, ikus daiteke eredu gehiago entrenatzeak, kasu honetan, orokorrean ez dituela ereduaren emaitzak hobetzen. Bestetik, “4” eta “5to4” zutabeetako emaitzak konparatuz, ikus daiteke hobekuntza bat dagoela “5to4”-ko emaitzetan. Hala ere, hori horrela da *bert-base-uncased* MedMCQA dataset-arekin fin-tzerakoan ‘5to4’ zutabeko emaitzak “4” zutabekoak baino hobekak direlako. Jarraitzeko,

*train* eta *dev* partizioak aldatuz lortutako emaitzak aztertuz, ez daude emaitzen artean desberdintasunik.

Gainera, *bert-base-uncased* MedMCQA *dataset*-arekin fintzerakoan lortutako emaitzak hobetzea ia ez da lortu eta lortzekotan 0,01 edo 0,08 baino ez da hobetu. Honek guztiak esan nahi du finduta dagoen modelo handi horrek ez duela ezer berririk ikasten CasiMédicos-MIR datu multzoa oso txikia delako. Izan ere, emaitzak hobetzea lortu du partizioetan dagoen datu kopurua handitu denean: *train* eta *dev* partizioetan 5 erantzun posible dituzten galderak moldatzerakoan (4. eta 5. errenkadak) eta *test* partizioan 5 erantzun posible dituzten galderak moldatzerakoan (“5to4” zutabea).

### Testuingurua erabiltzearen eragina

Testuingururekin lortutako emaitzak aztertuz, alde batetik ikus daiteke kasu honetan eredia gehiago entrenatzeak (konfigurazio normala aldatzeak) emaitzak hobetzen dituela. Beste alde batetik, “4” eta “5to4” zutabeetako emaitzak konparatuz, ikus daiteke “5to4”-ko emaitzak okerragoak direla. Hala ere, hori horrela da *bert-base-uncased* MedMCQA *dataset*-arekin fintzerakoan “5to4” zutabeetako emaitzak “4” zutabeetakoak baino okerragoak direlako. Jarraitzeko, *train* eta *dev* partizioak aldatuz lortutako emaitzak aztertuz, ez daude emaitzen artean ia desberdintasunik.

Bestalde, testuingururik gabeko egindako esperimentuetan bezala, *bert-base-uncased* MedMCQA *dataset*-arekin fintzerakoan lortutako emaitzak hobetzea ia ez da lortu. Eta hobetzea lortu da partizioetan dauden datu kopurua handitu denean.

### Testuingurua bai eta testuingurua ez erabiltzea

Laburbilduz, testuingurua erabilia edo erabili gabe *bert-base-uncased* MedMCQA *dataset*-arekin fintzerakoan lortutako emaitzak orokorrean hobetzea ez da lortu. CasiMédicos-MIR corpusetako datuak oso gutxi dira eredu handiak zerbait berri ikasteko.

#### 5.2.3.2 4 eta 5 aukera posible dituzten galderen esperimentuak

Azkenik, egin den beste esperimentu bat izan da MedMCQA datuekin findu egin den oinarritzko modeloari CasiMédicos-MIR corpus osoarekin fintzea. CasiMédicos-MIR corpusak 4 eta 5 erantzun posiblez osatuta dagoenez, probatu nahi izan da zein den ereduaren zehaztasuna 4 eta 5 erantzun posible dituzten galderetan. Hori probatzerakoan ereduaren mugak agerian geratu dira. Izan ere, ereduak bakarrik 4 erantzun posible dituzten galderen erantzun zuzena iragartzeko entrenatu da eta ez du uzten 5 erantzunekin probatzea.

#### 5.2.4 Bost erantzuneko eredu berria

Aurreko atalean esan bezala, oinarritzko ereduaren muga bat da ez duela 4 eta 5 erantzun posible dituzten galderen gainean funtzionatzen. Hori dela eta, 4 eta 5 erantzun posible dituzten galderetan erantzun egokia iragartzeko sortu den eredu berriarekin lortutako emaitzak 5.8. taulan agertzen dira. Eredua MedMCQA *dataset* aldatuaren gainean entrenatu da eta CasiMédicos-MIR corpus osoaren gainean eta CasiMédicos-MIR *test* partizioaren gainean probatu da. Taulako “5. hutsik” eta “5. EZ hutsik” zutabeek 5.6. taularen berdina adierazten dute.



Datu guztien gainean lortutako emaitzak txarragoak dira *test* partizioaren gainean lortutakoekin konparatuz. Hala ere, datu gehiago egonda ereduaren errealitateko errendimenduaren adierazpen egokiagoa islatzen dute.

Gainera, 4 erantzun posible dituzten kasuetan 5. aukera bat hutsik edo ez gehitzeak daukan eragina ezin da zehaztu. Testuinguruarekin emaitzak hobetu eta testuingururik gabe emaitzak okertzen batira.

Azkenik, testuingururik gabe lortzen diren emaitzak hobeak dira. Honen arrazoa egin den lehenengo esperimentuan, non eredu ere MedMCQA testuinguruarekin entrenatu den, aipatzen dena da. Eredua entrenatzeko erabili den *dataset*-ean dagoen testuinguru PubMed eta Wikipediatik sortua izan da. Baina ereduak horren gainean ikasten duenak ez dio CasiMédicos-MIR datuen gainean iragarpen egoki bat egiteko balio.

	w/o test		w test	
	5. hutsik	5. EZ hutsik	5. hutsik	5. EZ hutsik
Datu guztiak	0,27	0,26	0,24	0,24
Test partizioa	0,31	0,29	0,28	0,29

5.8 Taula: Sortutako eredu berriarekin lortutako zehaztasun emaitzak.

### 5.2.5 Ereduen arteko konparazioa

Egin diren esperimentu desberdinek agerian utzi dute oinarritzat hartu den ereduak bere mugak dituela, bakarrik 4 erantzun posible dituzten galderak erabili ahal direlako. Sortu den eredu berriarekin muga hori gainditzea lortu da. Horrekin lortutako emaitzak okerragoak dira baina, 5 erantzun asmatzearen ataza zailagoa da. Gainera, lortu diren emaitzak ausaz iragarri asmatzearen probabilitatea (%20) baino altuagoak dira.

Bestalde, 4 erantzunen kasuan, ikusi da CasiMédicos-MIR *corpusa* bakarrik erabiltzen denean emaitzak okerragoak direla *corpusa* txikia delako. Hala ere, CasiMédicos-MIR MedMCQA-rekin finduta dagoen eredu hobetzeko erabiltzen baldin bada, kasu gutxi batzuetan lortzen da zehaztasuna hobetzea. Izan ere, lanean [6] *bert-base-uncased*-ekin lortutako 0,37ko eta oinarritzko eredu CasiMédicos-MIR datuen gainean lortutako 0,44ko emaitza hobetzea lortu da. Esperimentu konkretu batean 0,56 lortu da, baina batezbestekoa kontuan izanik 0,47 lortu da. Aipatu behar da lanean medikuntzan espezializatuak dauden beste BERT batzuekin esperimentuak egin dituztela, emaitza hobeak lortuz. Baina lan honetan, *bert-base-uncased*-ekin emaitzak hobetzea helburua izan da.

Azkenik, CasiMédicos-MIR-en testuinguru erabiltzeak orokorrean emaitzak okertzen ditu, agerian utziz corpusaren testuinguruaren kalitatea. Bestetik, 4 erantzun posible dituzten galderetan 5. erantzun bat sartzen denean, emaitzak kontraesankorrak izan direnez, ezin izan da ondorioztatu zein den aukerarik egokiena: hura hutsik uztea edo ez uztea.

## 5.3 Bi hurbilpenen konparazioa

Bi hurbilpenak konparatzerako orduan MCQA hurbilpenak daukan beste muga bat agerian jartzen da. MCQA ingelesezko azterketekin baino ezin da probatu eta ezagutza-baseetan oinarritutako hurbilpenak, aldiz, eraikitako ezagutza-baseak daukan 5 hizkuntzetan (inge-

lesez, gaztelaniaz, euskaraz, frantsesez eta katalanez) dauden azterketekin proba daiteke. Hortaz, bakarrik ezagutza-baseetako hurbilpeneko ingelesezko emaitzekin konparatuko da.

Bi hurbilpenen emaitzak konparatzen baldin baditugu, MCQA-ren zehaztasuna altuagoa da. Ezagutza-baseetan oinarritutako hurbilpenak, aldiz, galdera gutxi ondo iragartzen ditu, baina kontuan izan behar da galdera gutxietan erantzun bat bueltatzeko gai dela. Beraz, erantzun bat bueltatzeko gai den galderen artetik zenbat asmatu dituen kontuan hartuz, zehaztasuna igotzen da.

Hori horrela, MCQA hurbilpena hobea da corpus osoa kontuan izaten badugu. Izan ere, hurbilpenak *test* tamaina desberdinekin ezagutza-baseetan oinarritutako hurbilpenak baino hobeto funtzionatzen du. Hala ere, ezagutza-base hurbilpenak erantzun bat bueltatzeko gai izan den kasuetan, MCQA hurbilpenaren errendimendua nolako den ikusi nahi izan da.

Ezagutza-base hurbilpenak iragarri dituen kasuak 4 eta 5 aukera posible dituztenez, kasu horiek MCQA hurbilpenean 4 eta 5 erantzun iragartzeko esperimenduetan probatu dira. Ezin da 5 erantzun dituzten galderak 4 aukera izatera moldatu ataza sinplifikatzen delako eta konparazioa bidezkoa ez delako. 5.9. taulan lortutako emaitzak ikus daitezke.

MCQA hurbilpenaren kasuan, probatu da oinarritzko ereduak erabiltzen duen *bert-base-uncased* CasiMédicos-MIR corpusarekin fintzea. Kasu honetan, CasiMédicos-MIR corpusa txikia denez *seed* desberdinekin probatu da eta bestetik, eredu gehiago entrenatzea probatu da. Lortutako emaitza guztien batezbestekoa eta desbiderapen estandarra kalkulatu da. Eredu berriak 4 eta 5 aukera dituzten galderekin erabili daitekeenez, hura ere probatu da. Kasu horretan, ereduak MedMCQA corpus handiaren gainean entrenatuta dagoenez ez da *seed* desberdinekin probatu eta ez da batezbestekoa eta desbiderapena kalkulatu, batezbesteko errenkadan lortutako zehaztasuna agertzen da. *Seed* desberdinekin lortutako emaitza bakoitza A. eranskinen aurki daiteke.

Emaitzak ikusirik bi hurbilpenen artean desberdintasun handirik ez dago. Ezagutza-base hurbilpenak “*bert-base-uncased* + CasiMédicos-MIR” esperimendua konfigurazio normalarekin lortutako emaitzak hobetzea lortzen du. Beste bi esperimenduekin konparatuz, aldiz, emaitza txarragoak lortu ditu. Hala ere, 0,06 eta 0,07-ko diferentzia baino ez dago. Gainera, ezagutza-base hurbilpenarekin desbiderapen estandarra txikiagoa da.

Laburbilduz, MCQA hurbilpenak orokorrean hobeto funtzionatzen du galdera guztiekin emaitza bat bueltatzeko gai delako. Baina ezagutza-base hurbilpenak gai izango balitz galdera guztiekin emaitza bat bueltatzen, seguruenik emaitza hobeak edo antzekoak lortuko lituzke, 5.9. taulako emaitzak ikustea baino ez dago. Bertan ikusten da ezagutza-baseak emaitza bat bueltatzeko gai den kasuetan MCQA hurbilpenaren antzeko emaitzak lortzen dituela. Gainera ezagutza-base hurbilpenak MCQA-ren mugak gainditzen ditu, hainbat hizkuntzetako eta edozein erantzun kopuruko galderekin erabili daiteke eta.

	Ezagutza-base hurbilpena	<i>bert-base-uncased</i> + CasiMédicos -MIR	Konfigurazio normala	epoch = 20 eta patience = 10	Eredu berria
Batezbestekoa	0,31	0,23	0,37	0,38	0,38
Desbiderapen estandarra	0,04	0,11	0,18	-	-

5.9 Taula: Ezagutza-base hurbilpenak emaitza bat bueltatu dituen galderak **bakarrik** kontuan harturik lortutako zehaztasuna esperimendu desberdinekin.

## Ondorioak eta etorkizunerako lana

Lan honetan egin diren ekarpen nagusiak hurrengoak dira:

- **Ezagutza-base eleanitz eta aberatsa sortzea.** Ezagutza-base eleanitz eta aberatsa sortu da beste ezagutza-baseen ezagutza bateratzen duena. Horretarako Wikidata, Wikipedia eta BabelNet erabili dira. Horrela, hizkuntza bakoitzeko 22.000 gaixotasun inguru dituen ezagutza-base aberats bat lortu da beste ezagutza-baseekin lotura ezartzen duena.
- **Multiple Choice Question Answering ataza ebaztea ezagutza-baseak erabiliz.** Eraikitako ezagutza-baseak izan ditzakeen aplikazio desberdinen artean, bertan dagoen informazioaz baliatuz MIR azterketen erantzun egokia iragartzeko erabili da. Horretarako, MER ataza egin da eta detektatutako gaixotasunen sintomak lortzeko eraikitako ezagutza-basea erabili da, UMLS kodea edota izena Levenshtein distantzia bidez bilatuz.

Emaitzek ezagutza-basearen estaldura falta agerian utzi dute, izan ere, sistema oso gutxitan aukera egokia iragartzeko gai izan da. Hala ere, iragartzeko gai izan den galderak bakarrik kontuan hartuz, emaitzek ausaz iragarri asmatzearen probabilitatea (%20) gainditzea lortu dute. Ondorioztatu dugu aplikazio honetarako eraikitako ezagutza-basea guztiz egokia ez dela. Hala ere, bere alde onak baditu; azterketak hainbat hizkuntzetan egon daitezke ezagutza-basea eleanitza delako eta bestalde, galderen aukera posible kopurua mugagabea izan daiteke. Baina geroz eta aukera gehiago izan, orduan eta zailagoa izango da.

- **Multiple Choice Question Answering ataza ebaztea hizkuntza-ereduak erabiliz.** Horrela, aurreko sistemaren errendimendua konparatu ahal izango da. Horretarako, [Pal et al.](#)-ren lana erreferentziatzen hartuta, dagozkion aldaketak egin zaizkio. Datuei dagokienez, lan horretan eraiki duten MedMCQA eta gure datuak (CasiMédicos-MIR) erabili dira.

Emaitzetan ikusi dugu, oinarriko eredu horrek bere mugak dituela. Alde batetik, bakarrik ingelesezko azterketekin proba daiteke eta beste alde batetik, bakarrik lau erantzun posible dituzten galderak erabili daitezke. Hala ere, azkenengo horri aurre egiteko beste eredu berri bat sortu dugu, zeinak ausaz iragarri asmatzearen

probabilitatea (%20) gainditzea lortzen baitu. Ereduaren mugak gainditzeaz gain, erreferentziako laneko emaitzak hobetzea lortu da 0,37tik 0,47ra. Gainera, bere eredu CasiMédicos-MIR datuen gainean lortutako 0,44ko emaitza ere hobetu da. Hala ere, erreferentziako lanean testuingurua erabiltzeak emaitzak hobetzen baditu ere, MIR azterketak dituen corpuseko (CasiMédicos-MIR) testuingurua erabiliz emaitzak ez dira hobetzen. Hori dela eta, ondorioztatu dugu CasiMédicos-MIR testuinguruaren kalitatea zalantzarria dela. Azkenik, ikusi da eredu CasiMédicos-MIR datuekin entrenatuz, egindako esperimentu desberdin batzuetan emaitzak hobetzen direla.

- **Bi hurbilpenak konparatu dira.** MCQA ataza ebazteko erabili diren bi hurbilpenak konparatu dira bakoitzak dituen abantailak eta desabantailak aztertzeko.

Emaitzek erakutsi dute hizkuntza-ereduak erabiliz egindako hurbilpenak galdera guztietan erantzun bat bueltatzen duenez, emaitzak hobeak direla. Ezagutza-baseak erabiliz egindako hurbilpenak, aldiz, ez da gai galdera guztietan erantzun bat bueltatzeko eta ondorioz, emaitzak txarragoak dira. Baina honek erantzun bat bueltatzen dituen galderak bakarrik kontuan harturik, biek antzeko emaitzak eskuratzen dituzte. Hori dela eta, ondorioztatu da ezagutza-baseak erabiliz gai izango balitz galdera guztiekin emaitza bat bueltatzen, hau da, bere estaldura handiago izango balitz, seguruenik emaitza hobeak edo antzekoak lortuko lituzkeela.

Laburbilduz, ikusi da MCQA ataza ebaztea erraza ez dela. Izan ere, egindako bi hurbilpenekin behin baino ez da %50eko zehaztasuna gainditu esperimentu konkretu batean. Hala ere, ezagutza-basea oso baliagarri eta aberatsa sortu da aplikazio desberdinak izan ditzakeenak etorkizunean. Gainera, gure sistemak oinarritzat hartzen duen ereduaren [6] emaitzak hobetzea lortu da eta honen mugak gainditu dira. Eraikitako eredu berri baten bidez 5 erantzun dituzten galderak ebaztea lortu da eta ezagutza-baseekin horretaz aparte, beste hizkuntza batzuetan idatzita dauden azterketak ebaztea lortu da.

## 6.1 Etorkizunerako lana

Lana garatu eta ebaluatu ostean, honek erakutsi dituen gabeziak hobetzeko asmoz, zenbait hobekuntza egin daitezke etorkizunean.

Hasteko, nahiz eta ezagutza-baseetan oinarritutako hurbilpenarekin lortutako emaitzak eraiki den ezagutza-basearen estalduraren ondorio izan, hobekuntza batzuk egin daitezke. Sistemaren parte garrantzitsu bat erantzunetan lortutako sintomak, galderen sintomekin komunean dituztenak detektatzea da. Hori hobetzeko, Levenshtein distantzia erabiltzeaz gain, entitateen desanbiguazioa, *entity linking* ingelesez, egin daiteke. Hori, Levenshtein ez bezala, antzekotasun semantikoan oinarritzen ohi da. Levenshtein neurriak hitzen azaleko formari erreparatzen dio, aldiz distantzia edo antzekotasun semantikoak hitzen esanahia hartzen du oinarri gisa. Horrela, adibidez, erantzun batean *migraña* agertuz gero eta galderetan dauden sintometan hori ez baldin badago, baina bai *dolor de cabeza* agertzen bada. Entitateen desanbiguazioaren bidez, galderako *dolor de cabeza* eta erantzuneko *migraña* berdinak direla ondorioztatuko lituzke semantika erabiltzen duelako lotura ezartzeko. Levenshtein distantziarekin, aldiz, ezinezkoa izango litzateke berdintzat kontsideratzea.

Jarraitzeko, hizkuntza-ereduen bidez egindako hurbilpena hobetzeko CasiMédicos-MIR corpusa handitu daiteke. Esperimentuen bidez ikusi baita datu kopurua urria dela

eta ereduaren eragina duela. Baita corpusaren testuingurua hobetu daiteke, testuingurua CasiMédicos-MIR webgunean mediku bolondresen idatzitako izan beharrean, [Pal et al.](#)-ren lanean erabiltzen dutena eraiki daiteke. Bertan *retriever+reader* pausoak erabili dituzte. Lehenik, galdera bakoitzeko, hura erantzuteko Wikipediako eta PubMed artikuluetako paragrafo garrantzitsuenak lortzen dira. Gero, galdera hobeto erantzuten duten paragrafoak hartzen dira eta horiekin eraikitzen den erantzun egokia testuinguru moduan erabiltzen da.

Azkenik, emaitzak hobetzeko bi hurbilpenak nahastu daitezke bakoitzaren abantailak aprobetxatuz. Horretako, bai hizkuntza-ereduak bai ezagutza-baseak baliatzen dituen sistema bat eraiki daiteke, “Knowledge infusion” [64, 65] deritzon teknikak erabiliz. Horretarako, sistema bat sortu daiteke eraikitako medikuntzarako baliabidearen informazioa (gaixotasunak, sintomak, diagnosia, kausak...) erabiliko duena. Ondoren, horrekin, hizkuntza-ereduen errendimendua hobetu daiteke. Gainera, erantzun egokia iragartzeaz gain, sistema erantzun egokia aukeratzearen azalpena emateko gai izango da. Izan ere, helburu berdina erdietsi duten lan batzuk aurki daitezke literaturan [65, 62].



# MCQA hurbilpeneko emaitzak

*Multiple Choice Question Answering* hurbilpenean esperimentu batzuk CasiMédicos-MIR corpus txikiaren gainean entrenatu eta ebaluatu direnez, bost *seed* desberdinekin probatu dira. *Seed* bakoitzarekin lortutako emaitzak atal honetan daude.

## A.1 Oinarrizko ereduaren *bert*-a CasiMédicos-MIR corpusarekin findu (*bert*-base-uncased + CasiMédicos-MIR )

### A.1.1 4 aukera posible dituzten galderen esperimentuak

Oinarrizko ereduari CasiMédicos-MIR corpusarekin finduz, 4 erantzun posible dituzten galderekin, lortutako esperimentuen emaitzak [A.1.](#) taulan agertzen dira.

### A.1.2 4 eta 5 aukera posible dituzten galderen esperimentuak

Oinarrizko ereduari CasiMédicos-MIR corpusarekin finduz, 4 eta 5 erantzun posible dituzten galderekin, lortutako esperimentuen emaitzak [A.2.](#) taulan agertzen dira.

## A.2 Hurbilpenen konparazioa

Bi hurbilpenak (ezagutza-baseak eta MCQA) konparatzeko egindako esperimentuen emaitzak [A.3.](#) taulan agertzen dira.

		w/o test		w test	
		4	5to4	4	5to4
Train eta dev-eko 4ko erantzunak, konfigurazio normala	seed 42	0,12	0,15	0,16	0,15
	seed 60	0,20	0,24	0,16	0,15
	seed 50	0,32	0,22	0,36	0,17
	seed 46	0,20	0,33	0,32	0,19
	seed 55	0,32	0,24	0,24	0,20
Batezbestekoa		<b>0,23</b>	<b>0,24</b>	<b>0,25</b>	<b>0,17</b>
Desbiderapen estandarra		<b>0,09</b>	<b>0,06</b>	<b>0,09</b>	<b>0,02</b>
Train eta dev-eko 4ko erantzunak, epoch=20, patience=10	seed 42	0,16	0,22	0,24	0,19
	seed 60	0,24	0,30	0,32	0,17
	seed 50	0,32	0,20	0,32	0,19
	seed 46	0,24	0,24	0,20	0,20
	seed 55	0,24	0,19	0,16	0,17
Batezbestekoa		<b>0,24</b>	<b>0,23</b>	<b>0,25</b>	<b>0,18</b>
Desbiderapen estandarra		<b>0,06</b>	<b>0,04</b>	<b>0,07</b>	<b>0,02</b>
Train eta dev-eko 4ko erantzunak eta 5koetan azkena kendu, konfigurazio normala	seed 42	0,12	0,33	0,20	0,15
	seed 60	0,24	0,24	0,12	0,20
	seed 50	0,40	0,26	0,32	0,14
	seed 46	0,16	0,30	0,28	0,09
	seed 55	0,12	0,20	0,28	0,26
Batezbestekoa		<b>0,21</b>	<b>0,27</b>	<b>0,24</b>	<b>0,17</b>
Desbiderapen estandarra		<b>0,12</b>	<b>0,05</b>	<b>0,08</b>	<b>0,06</b>
Train eta dev-eko 4ko erantzunak eta 5koetan azkena kendu, epoch=20, patience=10	seed 42	0,20	0,17	0,32	0,22
	seed 60	0,20	0,24	0,24	0,22
	seed 50	0,40	0,26	0,28	0,19
	seed 46	0,16	0,30	0,20	0,24
	seed 55	0,40	0,33	0,12	0,13
Batezbestekoa		<b>0,27</b>	<b>0,26</b>	<b>0,23</b>	<b>0,20</b>
Desbiderapen estandarra		<b>0,12</b>	<b>0,06</b>	<b>0,08</b>	<b>0,04</b>

**A.1 Taula:** Oinarrizko ereduaren *bert*-a CasiMédicos-MIR corpusarekin finduz (bert-base-uncased + CasiMédicos-MIR) lortutako emaitzak *seed* desberdinekin eta batezbestekoa eta desbiderapen estandarren kalkulua.



		w/o test		w test	
		5. hutsik	5. EZ hutsik	5. hutsik	5. EZ hutsik
Train eta dev osoa, konfigurazio normalarekin	seed 42	0,31	0,29	0,18	0,20
	seed 60	0,25	0,27	0,22	0,24
	seed 50	0,24	0,29	0,20	0,16
	seed 46	0,29	0,27	0,18	0,18
	seed 55	0,17	0,22	0,12	0,18
Batezbestekoa		<b>0,25</b>	<b>0,27</b>	<b>0,18</b>	<b>0,19</b>
Desbiderapen estandarra		<b>0,05</b>	<b>0,03</b>	<b>0,04</b>	<b>0,03</b>
Train eta dev osoa, epoch=20, patience=10	seed 42	0,27	0,25	0,24	0,24
	seed 60	0,31	0,16	0,16	0,25
	seed 50	0,27	0,25	0,20	0,22
	seed 46	0,24	0,20	<b>0,26</b>	0,22
	seed 55	0,22	0,25	0,20	0,37
Batezbestekoa		<b>0,26</b>	<b>0,22</b>	<b>0,21</b>	<b>0,26</b>
Desbiderapen estandarra		<b>0,03</b>	<b>0,04</b>	<b>0,04</b>	<b>0,07</b>

A.2 Taula: Oinarritzko ereduaren *bert*-a CasiMédicos-MIR corpus osoarekin, 4 eta 5 aukera dituzten galderekin, finduz lortutako emaitzak *seed* desberdinekin eta batezbestekoa eta desbiderapen estandarren kalkulua.

		Ezagutza-base hurbilpena	bert-base-uncased + Ca- siMédicos -MIR		Eredu berria
			Konfigurazio normala	epoch = 20 eta patience = 10	
Emaitza bat bueltatutako guztiak	seed 42	0,29	0,13	0,31	-
	seed 60	0,35	0,19	0,69	-
	seed 50	0,32	0,38	0,31	-
	seed 46	0,32	0,13	0,25	-
	seed 55	0,26	0,31	0,31	-
Batezbestekoa		<b>0,31</b>	<b>0,23</b>	<b>0,37</b>	<b>0,38</b>
Desbiderapen estandarra		<b>0,04</b>	<b>0,11</b>	<b>0,18</b>	-

A.3 Taula: Ezagutza-base hurbilpenak erantzun bat bueltatutako galderak bakarrik kontuan harturik lortutako zehaztasuna esperimendu desberdinekin.



# Bibliografia

- [1] H. Turki, T. Shafee, M. A. H. Taieb, M. B. Aouicha, D. Vrandečić, D. Das, and H. Hamdi, “Wikidata: a large-scale collaborative ontological medical database,” *Journal of Biomedical Informatics*, vol. 99, p. 103292, 2019. Ikusi [v, 17, 21, and 28](#) orrialdeak.
- [2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.Ñ. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017. Ikusi [v, 18](#) orrialdeak.
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018. Ikusi [v, 19](#) orrialdeak.
- [4] D. Jin, S. Gao, J.-Y. Kao, T. Chung, and D. Hakkani-tur, “Mmm: Multi-stage multi-task learning for multi-choice reading comprehension,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 8010–8017. Ikusi [v, 20, 21, and 24](#) orrialdeak.
- [5] A.Ñixon, L. Fang, J. M. Havrilla, and K. Wang, “Termviewer—a web application for streamlined human phenotype ontology (hpo) tagging and document annotation,” *Chemistry & Biodiversity*, vol. 19, no. 12, p. e202200805, 2022. Ikusi [v, 23](#) orrialdeak.
- [6] A. Pal, L. K. Umapathi, and M. Sankarasubbu, “Medmcqa: A large-scale multi-subject multi-choice dataset for medical domain question answering,” in *Conference on Health, Inference, and Learning*. PMLR, 2022, pp. 248–260. Ikusi [v, 3, 22, 24, 26, 34, 42, 44, 45, 50, 55, 57, 61, 63, 64, and 65](#) orrialdeak.
- [7] Y. Mintz and R. Brodie, “Introduction to artificial intelligence in medicine,” *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 2, pp. 73–81, 2019. Ikusi [1](#) orrialdea.
- [8] D. Vilares and C. Gómez-Rodríguez, “Head-qa: A healthcare dataset for complex reasoning,” *arXiv preprint arXiv:1906.04701*, 2019. Ikusi [2, 24, and 34](#) orrialdeak.
- [9] J. McCarthy, “What is artificial intelligence,” 2007. Ikusi [13](#) orrialdea.
- [10] K. Chowdhary and K. Chowdhary, “Natural language processing,” *Fundamentals of artificial intelligence*, pp. 603–649, 2020. Ikusi [13](#) orrialdea.
- [11] O. Bodenreider, “The unified medical language system (umls): integrating biomedical terminology,” *Nucleic acids research*, vol. 32, no. suppl\_1, pp. D267–D270, 2004. Ikusi [16](#) orrialdea.
- [12] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. Ikusi [18, 24](#) orrialdeak.
- [13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. Ikusi [18, 24](#) orrialdeak.
- [14] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, “Aligning books and movies: Towards story-like visual explanations by watching movies and reading books,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27. Ikusi [19](#) orrialdea.

- [15] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, “Google’s neural machine translation system: Bridging the gap between human and machine translation,” *arXiv preprint arXiv:1609.08144*, 2016. Ikusi 19 orrialdea.
- [16] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” *arXiv preprint arXiv:1802.05365*, 2018. [Online]. Available: <https://arxiv.org/abs/1802.05365> Ikusi 19 orrialdea.
- [17] M. Marrero, J. Urbano, S. Sánchez-Cuadrado, J. Morato, and J. M. Gómez-Berbis, “Named entity recognition: fallacies, challenges and opportunities,” *Computer Standards & Interfaces*, vol. 35, no. 5, pp. 482–489, 2013. Ikusi 20 orrialdea.
- [18] D. Jurafsky and J. H. Martin, *Speech and Language Processing (3rd ed. draft)*. Draft available online, 2018. Ikusi 20 orrialdea.
- [19] L. A. Ramshaw and M. P. Marcus, “Text chunking using transformation-based learning,” *Natural language processing using very large corpora*, pp. 157–176, 1999. Ikusi 20 orrialdea.
- [20] Y. Liu, S. Chowdhury, C. Zhang, C. Caragea, and P. S. Yu, “Interpretable multi-step reasoning with knowledge extraction on complex healthcare question answering,” *arXiv preprint arXiv:2008.02434*, 2020. Ikusi 21, 24 orrialdeak.
- [21] S. Marro, B. Molinet, E. Cabrio, and S. Villata, “Natural language explanatory arguments for correct and incorrect diagnoses of clinical cases,” in *ICAART 2023-15th International Conference on Agents and Artificial Intelligence*, vol. 1, 2023, pp. 438–449. Ikusi 21 orrialdea.
- [22] D. Diomedi and A. Hogan, “Question answering over knowledge graphs with neural machine translation and entity linking,” *arXiv preprint arXiv:2107.02865*, 2021. Ikusi 22 orrialdea.
- [23] K. Chard, M. Russell, Y. A. Lussier, E. A. Mendonça, and J. C. Silverstein, “A cloud-based approach to medical nlp,” in *AMIA Annual Symposium proceedings*, vol. 2011. American Medical Informatics Association, 2011, p. 207. Ikusi 22 orrialdea.
- [24] A. Pesaranhader, S. Matwin, M. Sokolova, and A. Pesaranhader, “deepbiowds: effective deep neural word sense disambiguation of biomedical text data,” *Journal of the American Medical Informatics Association*, vol. 26, no. 5, pp. 438–446, 2019. Ikusi 22 orrialdea.
- [25] H. Liu, Y. A. Lussier, and C. Friedman, “Disambiguating ambiguous biomedical terms in biomedical narrative text: an unsupervised method,” *Journal of biomedical informatics*, vol. 34, no. 4, pp. 249–261, 2001. Ikusi 22 orrialdea.
- [26] Z. Liu, M. Yang, X. Wang, Q. Chen, B. Tang, Z. Wang, and H. Xu, “Entity recognition from clinical texts via recurrent neural network,” *BMC medical informatics and decision making*, vol. 17, pp. 53–61, 2017. Ikusi 22 orrialdea.
- [27] T. Munkhdalai, F. Liu, H. Yu *et al.*, “Clinical relation extraction toward drug safety surveillance using electronic health record narratives: classical learning versus deep learning,” *JMIR public health and surveillance*, vol. 4, no. 2, p. e9361, 2018. Ikusi 22 orrialdea.
- [28] Z. Li, Z. Yang, C. Shen, J. Xu, Y. Zhang, and H. Xu, “Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text,” *BMC medical informatics and decision making*, vol. 19, pp. 1–8, 2019. Ikusi 22 orrialdea.
- [29] B. Pandey, D. K. Pandey, B. P. Mishra, and W. Rhmann, “A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5083–5099, 2022. Ikusi 22 orrialdea.
- [30] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, “Biobert: a pre-trained biomedical language representation model for biomedical text mining,” *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020. Ikusi 22, 23 orrialdeak.
- [31] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon, “Domain-specific language model pretraining for biomedical natural language processing”

- ACM Transactions on Computing for Healthcare (HEALTH)*, vol. 3, no. 1, pp. 1–23, 2021. Ikusi 22 orrialdea.
- [32] E. Alsentzer, J. R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, and M. McDermott, “Publicly available clinical bert embeddings,” *arXiv preprint arXiv:1904.03323*, 2019. Ikusi 22 orrialdea.
- [33] P. Ontalvilla, A. Atutxa, and M. Oronoz, “Osasun-arloko entitate izendunen etiketatzea,” *V. Iker gazte. Nazioarteko ikerketa euskaraz*, p. 154, 2023. Ikusi 22, 23 orrialdeak.
- [34] R. Grishman and B. M. Sundheim, “Message understanding conference-6: A brief history,” in *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 1996. Ikusi 22 orrialdea.
- [35] E. F. Sang and F. De Meulder, “Introduction to the conll-2003 shared task: Language-independent named entity recognition,” *arXiv preprint cs/0306050*, 2003. Ikusi 22 orrialdea.
- [36] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. M. Strassel, and R. M. Weischedel, “The automatic content extraction (ace) program-tasks, data, and evaluation.” in *Lrec*. Lisbon, 2004, pp. 837–840. Ikusi 22 orrialdea.
- [37] G. Demartini, T. Iofciu, and A. P. De Vries, “Overview of the inex 2009 entity ranking track,” in *Focused Retrieval and Evaluation: 8th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2009, Brisbane, Australia, December 7-9, 2009, Revised and Selected Papers 8*. Springer, 2010, pp. 254–264. Ikusi 22 orrialdea.
- [38] K. Balog, P. Serdyukov, and A. P. d. Vries, “Overview of the trec 2010 entity track,” Norwegian univ of Science and Technology Trondheim, Tech. Rep., 2010. Ikusi 22 orrialdea.
- [39] S. R. Kundeti, J. Vijayananda, S. Mujjiga, and M. Kalyan, “Clinical named entity recognition: Challenges and opportunities,” in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 1937–1945. Ikusi 23 orrialdea.
- [40] H. Zhu, I. C. Paschalidis, and A. Tahmasebi, “Clinical concept extraction with contextual word embedding,” *arXiv preprint arXiv:1810.10566*, 2018. Ikusi 23 orrialdea.
- [41] Y. Si, J. Wang, H. Xu, and K. Roberts, “Enhancing clinical concept extraction with contextual embeddings,” *Journal of the American Medical Informatics Association*, vol. 26, no. 11, pp. 1297–1304, 2019. Ikusi 23 orrialdea.
- [42] A. R. Aronson, “Effective mapping of biomedical text to the umls metathesaurus: the metamap program.” in *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001, p. 17. Ikusi 23 orrialdea.
- [43] G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, “Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications,” *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507–513, 2010. Ikusi 23 orrialdea.
- [44] A. Casillas, N. Ezeiza, I. Goenaga, A. Perez, and X. Soto, “Measuring the effect of different types of unsupervised word representations on medical named entity recognition,” *International journal of medical informatics*, vol. 129, pp. 100–106, 2019. Ikusi 23 orrialdea.
- [45] A. Chaturvedi, O. Pandit, and U. Garain, “Cnn for text-based multiple choice question answering,” in *ACL 2018-56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 272–277. Ikusi 24 orrialdea.
- [46] S. Roy, N. Ehtesham, M. S. Islam *et al.*, “Augmenting bert with cnn for multiple choice question answering,” in *2021 24th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2021, pp. 1–5. Ikusi 24 orrialdea.
- [47] M. Tan, C. d. Santos, B. Xiang, and B. Zhou, “Lstm-based deep learning models for non-factoid answer selection,” *arXiv preprint arXiv:1511.04108*, 2015. Ikusi 24 orrialdea.

- [48] M. Yasunaga, H. Ren, A. Bosselut, P. Liang, and J. Leskovec, “Qa-gnn: Reasoning with language models and knowledge graphs for question answering,” *arXiv preprint arXiv:2104.06378*, 2021. Ikusi 24 orrialdea.
- [49] J. Bao, N. Duan, Z. Yan, M. Zhou, and T. Zhao, “Constraint-based question answering with knowledge graph,” in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers*, 2016, pp. 2503–2514. Ikusi 24 orrialdea.
- [50] T.Ñ. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016. Ikusi 24 orrialdea.
- [51] H. Kim and P. Fung, “Learning to classify the wrong answers for multiple choice question answering (student abstract),” in *Proceedings of the AAAI conference on artificial intelligence*, 2020, pp. 13 843–13 844. Ikusi 24 orrialdea.
- [52] P. Lu, S. Mishra, T. Xia, L. Qiu, K.-W. Chang, S.-C. Zhu, O. Tafjord, P. Clark, and A. Kalyan, “Learn to explain: Multimodal reasoning via thought chains for science question answering,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 2507–2521, 2022. Ikusi 24 orrialdea.
- [53] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019. Ikusi 24 orrialdea.
- [54] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “Squad: 100,000+ questions for machine comprehension of text,” *arXiv preprint arXiv:1606.05250*, 2016. Ikusi 24 orrialdea.
- [55] G. Lai, Q. Xie, H. Liu, Y. Yang, and E. Hovy, “Race: Large-scale reading comprehension dataset from examinations,” *arXiv preprint arXiv:1704.04683*, 2017. Ikusi 24 orrialdea.
- [56] D. Jin, E. Pan, N. Oufattole, W.-H. Weng, H. Fang, and P. Szolovits, “What disease does this patient have? a large-scale open domain question answering dataset from medical exams,” *Applied Sciences*, p. 6421, 2021. Ikusi 24, 34 orrialdeak.
- [57] J. Li, S. Zhong, and K. Chen, “Mlec-qa: A chinese multi-choice biomedical question answering dataset,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 8862–8874. Ikusi 24 orrialdea.
- [58] Y. Labrak, A. Bazoge, R. Dufour, M. Rouvier, E. Morin, B. Daille, and P.-A. Gourraud, “French-medmcqa: A french multiple-choice question answering dataset for medical domain,” *arXiv preprint arXiv:2304.04280*, 2023. Ikusi 24, 34 orrialdeak.
- [59] Q. Jin, B. Dhingra, Z. Liu, W. W. Cohen, and X. Lu, “Pubmedqa: A dataset for biomedical research question answering,” *arXiv preprint arXiv:1909.06146*, 2019. Ikusi 24, 34 orrialdeak.
- [60] A. Pampari, P. Raghavan, J. Liang, and J. Peng, “emrqa: A large corpus for question answering on electronic medical records,” *arXiv preprint arXiv:1809.00732*, 2018. Ikusi 24 orrialdea.
- [61] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, “Measuring massive multitask language understanding,” *arXiv preprint arXiv:2009.03300*, 2020. Ikusi 24 orrialdea.
- [62] D. Sileo, K. Uma, and M.-F. Moens, “Generating multiple-choice questions for medical question answering with distractors and cue-masking,” *arXiv preprint arXiv:2303.07069*, 2023. Ikusi 24, 65 orrialdeak.
- [63] G. Navarro, “A guided tour to approximate string matching,” *ACM computing surveys (CSUR)*, vol. 33, no. 1, pp. 31–88, 2001. Ikusi 38 orrialdea.
- [64] K. Jha and A. Zhang, “Continual knowledge infusion into pre-trained biomedical language models,” *Bioinformatics*, vol. 38, no. 2, pp. 494–502, 2022. Ikusi 65 orrialdea.
- [65] Y. He, Z. Zhu, Y. Zhang, Q. Chen, and J. Caverlee, “Infusing disease knowledge into bert for health question answering, medical inference and disease name recognition,” *arXiv preprint arXiv:2010.03746*, 2020. Ikusi 65 orrialdea.