



DONOSTIA / SAN SEBASTIÁN 2016
EUROPAKO KULTUR HIRIBURUA



LEHENENGO EGUNA

[codefest](#) • [uztaila 5, 2016](#)

Astelehenean eman zaio hasiera CODEFEST 2016 udako eskolari. Gosari oparo baten ostean, proiektuak proposatzeko aukera egon da, baita proiektuetara batzeko ere. Programazioaren eta hizkuntzalaritzaren inguruko ezagutzak uztartzeko aukera izango dute CODEFESTeko partaide ezberdinek, proiektu berriak garatzeko, sortzeko, elkartrukatzeko, laguntzeko etab. Bestalde, proiektuok euren hizkuntzekin aberasteko edota euren hizkuntzetara ekartzeko aukera izango dute aste honetan zehar.

Hizkuntza anitzeko galdera-erantzunak erronka jokoetarako Wikipediarekin

Lehenengo egunean taldekideak quiz eleanitzaren lehen pausoak garatzen aritu dira: kultura orokorraren inguruko Wikipedia artikuluak edota pasarteak itzuli eta egokitu dituzte. Ingelesetik euskarara itzuli dituzte. Horrez gain, programa berezitu bat baliatu dute euskaraz idatzitako Wikipedia artikuluak jasotzeko, aurreprozesatzeko, garbitzeko eta tokenizatzeko (hitzetan segmentatzea).

Hizkuntza gutxituarentzako morfologia-analizatzailea sortzea

Mans Huldenen ardurapean hizkuntza gutxituentzako morfologia-analizatzaile bat sortzeko asmoz batu dira taldera, besteak beste, Elena Badmaeva (Donostian hizkuntza-gaitasunaren inguruko masterra) eta Chenchen Julio Song (Cambridge unibertsitatean sintaxi eta morfologia teorikoaren inguruko tesia). Wiktionary bezalako tresnari esker hungariarraren datu-base lexikala sortzen ari dira. Foma kode irekiko finite-state tresna multzoa erabiliz eta hungariarren inguruan Juliok duen ezagutzarekin uztartuz, morfologia-analizatzaile ireki bat sortu nahi dute hungariarrantzat.

HAP literatura aztertu eta sortzeko

Taldearen lana literatura testuak prozesatzeko tresna bat sortzea izan da. Horretarako, haur literaturari lotutako ipuinak prozesatu dituzte; kasu honetan, Txanogorritxu izan da aukeratutakoa. Besteak beste, ipuinean erabilitako hizkuntza aldatu da; adjektiboak aldatu dira pertsonaien jarrera aldatzeko; gertakizunen ordena aldatu da... Horrez gain, beste hainbat hizkuntza eta hizkera nahastu dira ipuina prozesatzean.

Behagunea

Sei pertsona ari dira taldean parte hartzen [Behagunea](#) proiektua garatzeko: Iñaki (Elhuyarreko partaidea), Viktor (Errumaniakoa), Dani (katalanarekin lan egiten du), Cassandra (ingenieria, Greziatik dator), Sabrina (masterra informatikan egiten ari da, Italiatik dator) eta Rodrigo (Ixa taldeko kidea). Guztiak daude prest hizkuntzen arteko lana eta partaidetza sustatzeko (katalana, italiara, eta batez ere taldekideen artean dauden

hizlarien hizkuntzak edo ama-hizkuntzak). Horrez gain, hiztegia edo datuak sustatzeko baliabideak eta euren erabilera hobetzea eta hedatzea dute helburu.

Taldean "nahasten" diren hizkuntzen hiztegi bat prestatu nahi da, besteak beste, katalanarena, italiararena eta grezierarena, eta gero, hiztegi hori igoko da CODEFEST webgunera. Guztiak hasi dira lanean hiztegi hori osatzeko eta kasu askotan, Facebook eta Twitter sareak erabili dira lana errazteko.

ZureTTS plataforma beste hizkuntza batzuekin aberastea

CODEFEST 2016 udako laborategian, [ZureTTS](#) tresna beste hizkuntza batzuetan erabilgarri jartzeko proiektua aurkeztu dute Aholab taldeko kideek. ZureTTS proiektuaren webguneak erabiltzaileen ahotsaren bertsio sintetikoa sortzen du, grabaketetan oinarrituta. Ahots gaelikoa emateko irrikaz, Cathal Peelo (hizkuntzalaria) Geraldine Uí Bhriain (Gaeliko itzultzailea) eta Séamus Ó Briain (Gaeliko eta IKTetan aditua) hurbildu dira lantaldera, besteak beste. ZureTTS tresnari esker, adibidez, endekapenezko ahots-patologiak dituzten edo laringektomiak programatuta dituzten pertsonentzako komunikazio-tresna pertsonalizatuak sortu daitezke. Hain zuzen ere, Bríd Joe Jack abeslari irlandarraren egoerak piztu du hirukote irlandarraren interesa eta gaelikoz mintzatuko den ahots sintetikoa sortu nahi dute ZureTTS tresnarentzat.

Bestalde, Aholab-eko kideek (Inma Hernaez, Eva Navas, Daniel Erro, Ibon Saratxaga) iparraldeko errealitatera gerturatu nahi dute ZureTTS tresna. Hortaz, Iparrahotsa proiektua garatzen jarraituko dute, lapurteraren bertsio bat martxan jartzeko asmoz. ZureTTS Android aplikaziori dagokionez, Adam Varga arituko egokitzapen eta hobekuntza lanetan.

SRL eta Dockers

Proiektu paralelo ezberdinak garatzen dihardute lan talde honetan. Batetik Ixa taldeko zenbait partaidek testu multzo handi bat tratatzeko tresna linguistiko bat garatu nahi dute, SRL (semantic role labelling) edo etiketatze-prozesatzaile bat lortzeko asmoz. Bestetik, Eleanor Duttonek (psikolinguista) hizkuntza prozesamenduaren inguruan ari da lanean eta Suhail Sarwan (e-learning: ikasketa eta teknologia berriak uztartzen ditu) semantikaren arloan lanean dihardu.

Hizkuntza gutxituentzako itzulpen automatikoa

Francis Tyers eta Nora Aranberrik zuzendutako talde honen helburua hizkuntza gutxituentzako itzulpen automatikorako herramintak sortzea da. Herraminta horiek testu paraleloen korpusean oinarritzen diren algoritmoak erabiltzen dituzte eta hizkuntza gutxituek korpus urriak dituztenez, bestelako baliabideak erabili behar dituzte: gramatika-etiketazioa, analisi sintaktikorako programak, etab.

BIGARREN EGUNA

[codefest](#) • [uztaila 6, 2016](#)

Bigarren egunean, proiektu ezberdinetako taldekideengana jo dugu, euren lanaren nondik norakoak eta aurrerapausoak gertutik ezagutzeko. Zenbait taldekide ezagutzeko aukera izan dugu ere.

Hizkuntza anitzeko galdera-erantzunak erronka jokoetarako Wikipediarekin

Galdera-erantzunetan oinarritutako quiz proiektua aurrera doa. Lehenengo prototipoa martxan jartzeko asmoz, astelehenean bildutako eta itzulitako galdera sortak baliatu nahi dituzte. Prototipo honek euskal Wikipediako artikuluak quiz-eko pistekin erlazionatzen ditu eta pista gehienekin bat datorren artikulua atzematen du.

Hizkuntza gutxituarentzako morfologia-analizatzailea sortzea

Hungariararentzat morfologia analizatzaile bat garatzeko Ixa taldeko Iñaki Alegria eta Montse Maritxalar batu dira taldera, programazio lanetan laguntzeko asmoz. Hungariararen erro lexikalen zerrenda osatu ostean, besteak beste, izen, aditz eta adjetiboen araberako sailkapena egin dute. Gerora sailkapen hori programa berezi baten bidez konputatu nahi dute lexc formatuan.

HAP literatura aztertu eta sortzeko

[Story buffet](#) izendatu dute literatura analisirako eta sorkuntzarako garatu nahi duten tresna. Hizkuntzalariak, programatzaileak eta euren burua "hibrido" gisa dute hainbat aditu dihardute proiektuan lanean. Bigarren egunean, Ixa taldeko kideen (proiektu honen arduradunak) lana ezagutzeko parada izan dugu. Manex Agirrezabal poesiaren analisi metrikoan aditua da; programazioaren inguruan duen ezagutza dela medio, ipuinak arlo semantikotik aldatzeko aukera interesgarria Iruditzen zaio. Itziar Gonzalez hizkuntzalaria da jatorriz, baina programazioaren ezagutza barneratu izan du azken urteetan zehar; testuen konplexutasunaren analisisian eta sinplifikazioan interesatua dago.

Behagunea

[Behagune](#) proiektua garatzen jarraitu dute bigarren egunean zehar, taldekideen lan eta ezagutza ezberdinak uztartuz. Viktorrek (programatzailea) Ixa-pipes bisualizazioaren emaitzak aztertu ditu eta interfazeari itxura atsegin bat emateko ari da lanean. Bestalde, Danik (informatikaria) Ixa-pipes baliabideak katalanez erabiltzeko garatu nahi ditu. Sabrinak (hizkuntzalaria), Iñakiren (programatzailea) laguntzaz, sentimendu-analizatzaile bat jarri du martxan Behagune proiektuan oinarrituta. Amaitzeko, Kassandrak alde batera utzi du Behagune web-orrian sare sozialak txertatzeko proiektua, arazotsua suertatu baita. Eurokopa futbol txapelketaren harira idatzitako tweet-ak behatzeari ekin dio horren ordez.

ZureTTS plataforma beste hizkuntza batzuekin aberastea

[ZureTTS](#) proiektuaren hainbat aspektu garatzen ari dira. Batetik Aholab-eko taldekideak Iparrahotsa bertsioaren garapenean zentratu dira eta ahots emaleentzako galdera sortak zein interfazea prestatzen aritu dira. ZureTTSaren android aplikazioari dagokionez, ordea, akatsak identifikatzen eta plataforma instalatzeko prestaketak egiten jardun dute eguna. Amaitzeko, irlandako taldekideek web orriaren interfazea gaelikora itzuli dute eta kontaktuan jarri dira irlandako zenbait aditu eta unibertsitateekin, datu-base egoki bat eskuratzeko asmoz.

SRL eta Dockers

SRL proiekturako egituratu gabe dagoen datu-base bat egituratzen aritu dira, gerora ohar egokiak egiteko eta manipulatu ahal izateko. Suhail Sarwanek adierazi duenez SRL-en garapenak zuzeneko onura ekarri dezake semantiken arloan, batez ere e-learning gisako ikasketa ereduak sustatu eta hobetu nahi baldin bada. Besteak beste, Rodrigo Agerriren laguntzaz landu dituzte SRL-ak eta Eleanor Duttonek hizkuntza analisirako tresna baten garapenean txertatu nahi ditu, marokoko arabierarekin erabili ahal izateko, hain zuzen ere.

Hizkuntza gutxituentzako itzulpen automatikoa

Taldekide bakoitza ezagutzen duen hizkuntza konbinazio batetan zentratu da. Adibidez, Apertium sistema erabiliz frantses-okzitaniera hizkuntza konbinazioaz lanean hasi dira, okzitanierarentzat hizkuntza-analizatzaile bat garatzeko asmoz. Sistema berberaz Tetun-portuges (Timor orientaleko hizkuntza ofizialak) garatzeari ekin diote.

Beste taldekide batzuk, Matxin tresnari esker, euskara-gaztelera hizkuntza konbinazioen arteko transferentzian lexikoa prestatzen hasi dira. Programa horren bidez ere ingelesa-galesa itzultzailea garatu dute eta ingeles-eskara itzultzailea hobetzen aritu dira (dagoeneko erabilgarria baina akatsekin). Ingeles-euskara itzultzailea medikuntza alorrerako moldatuko dute ere.

HIRUGARREN EGUNA

[codefest](#) • [uztaila 7, 2016](#)

Asteazkenean lanean hasi aurretik, **Wilco van Duinkerke** (Amsterdameko Trivagoko software ingeniaria) bere esperientzia partekatu zuen Codefest-en. HNP teknologiek (hizkuntza naturalaren prozesamendua) Trivago proiektuan dute garrantzia azaldu zuen eta etorkizunean garatu nahiko lituzketen proiektuen berri eman zuen. Trivagoko beste kide batzuk ere ari dira parte hartzen CODEFESTeko proiektu ezberdinetan, hala nola, Min Fang, Cassandra Charalampidou eta Victor Blaga.

Gaurkoan bestalde, CODEFESTeko zenbait partaide eta antolatzaileekin hitz egiteko aukera izan dugu. Udako eskolaren xede eta baloreak aurkeztu dizkigute, eta etorkizunari begira egitasmo honek izan ditzakeen aukeren inguruan mintzatu dira.

Solasean...



ELIZABETE MANTEROLAekin

EHUko Letren fakultateko irakaslea da eta CODEFEST 2016 talde-eragileko kidea da. Hizkuntza gutxituak teknologien garapenarekin uztartuko dira udako eskola honetan. Itzulpengintzaren eta hizkuntzalaritzaren alorretik ekarpena egiteko asmoz batu da Elizabete egitasmo honetara.

“Kode irekiko proiektu ezberdinak garatu, sortu edota partekatu nahi dira aste honetan zehar eta beste hizkuntzentzat zein hizkuntzetan eskuragarri jarri nahi dira ere. Hortaz, eta hizkuntza gutxituek baliabide urriak izaten dituztela kontuan izanik, ezinbestekoa da produktuok libreki erabili eta garatzeko aukera izatea. Bestalde, CODEFESTen helburu nagusien artean elkarlana eta elkartrukea daude: hizkuntzen artean, teknologia garatzaile eta eragile ezberdinen artean, informatikarien eta hizkuntzalarien artean etab.”



FRANCIS TYERekin

Postdoktoretza ikerkertzalea da eta CODEFESTen itzulpen automatikoaren eta analisi morfologikoaren inguruan lanetan diharduten bi talderen tutorea da.

“Erabiltzen ari garen metodologia partaideen autonomian oinarrituta dago, nik lagundu eta gidatu baino ez dut egiten. Produktu ezberdinetan ari gara lanean: alde batetik, okzitaniera-frantsesa, tetum-portugesa eta galesa-ingelesa itzultzaile automatikoak sortzen ari gara; bestetik, Matxin izeneko euskara-ingelesa-gaztelera itzultzaile automatikoa garatu nahi dugu. Astebeteko udako eskola hau Hackathon gisa izendatu dezakegu. Ez dago lan egiteko formula zehatzik, toki bereberan batzen den talde bat da. CODEFESTek nahien eta beharren arabera ezagutza elkartrukatzeko eta disziplinen artean lan egiteko aukera paregabea eskaintzen du.”



RODRIGO AGERRIrekin

IXA taldeko ikertzailea da eta CODEFESTen talde-eragilean parte hartzeaz gain tutorea da 3 talde ezberdinetan.

“Hainbat proiektutan parte hartzeko eta tutorizatzeko aukera izan dut egun hauetan zehar, adibidez, Behagunean. Donostia 2016 barruan kokatzen den proiektua da Behagunea eta otsailean jarri zen martxan Ixa taldearen eta Elhuyar Fundazioaren arteko elkarlanari esker. Internet eta sare sozialen azterketan oinarritzen den proiektu bat da eta Donosti 2016ko testuinguruan gertatzen den gertaera garrantzitsuen inguruko iritziak, komentarioak, txioak eta bestelakoak jaso eta analizatzeko xedea du. CODEFEST udako eskola Behagunean garatzeko aukera paregabea iruditu zitzaigun. Dagoeneko lau hizkuntzatan (euskara, ingelesa, frantsesa eta gaztelera) baliagarri egon arren, bestelako hizkuntzak gehituko dira aste honetan zehar proiektura, adibidez, katalana eta italiarra.”

LAUGARREN EGUNA

[codefest](#) • [uztaila 7, 2016](#)

Martxan jarri aurretik, **Claudia Soria** ikerlariak (Pisako Hizkuntzalaritza Konputazionalaren Institutua, Italia) koordinatzen duen Erasmus+ proiektua aurkeztu digu: “[The Digital Language Diversity Project](#)”. Mundu digitalean Europako hizkuntza gutxituak edota nagusiak ez diren hizkuntzen presentzia aztertu nahi dute. Hizkuntza ezberdinen egoera eta sareko-bizitasuna ebaluatzeko galdetegi bat prestatu dute. Elhuyarrekin elkarlanean adibidez, euskara hizkuntza digitala den aztertzeko [galdera-sorta](#) bat prestatu dute. Bestek beste, jasotako informazioetik abiatuta eta komunitate ireki baten ideologiari jarraiki, hizkuntza gutxituak sarean biziberritzeko neurriak eta programa bat jarri nahi dute martxan.

Gaurkoan ere, proiektuetan lan egiteko azken eguna dela aprobetxatuz, CODEFESTeko partaide zein antolatzaileengana hurreratu gara.

Solasean...



EVA NAVASekin

Aholab taldeko ikerlaria da eta EHUko irakasle. Ahots sintetikoa garatzeko [ZureTTS](#) proiektuan parte hartzen ari da.

“Hizkuntza naturalaren prozesamendua (HNP), oro har, testuetan oinarritzen da. Aholab-en ordea ahots seinalea lantzen dugu. Ebakuntza baten ondorioz, adibidez, ahotsa galdu behar baduzu, aurretik zure ahotsa grabatu dezakezu, gerora zure ahots sintetiko propioa izateko. ZureTTS Android aplikazioa mugikorrean funtzionatzen du. Momentuz euskaraz, gaztelera, galegoz, katalanez eta ingelesez dago eskuragarri eta laster gaelikoz eta lapurteraz ere izango da. Etorkezunari begira hizkuntza gehiago txertatu nahiko genituzke sisteman, baina egun batez ere, [ahots-emaleen bila](#) gabiltza.”



AITZOL ASTIGARRAGArekin

Donostia 2016 proiektu digitalen arduraduna da eta CODEFESTen talde-eragileko partaidea.

“Espektatiba orokorrak bete ditugu. Kuantitatiboki 55 partaide eta 24 hizkuntza ezberdin bildu gara, eta 8 proiektu ezberdin gauzatu ari dira. Kualitatiboki hitz egiteko, ostera, ostiraleko proiektuen aurkezpena arte itxaron beharko dugu, baina ziur emaitza interesgarria izango dela. Bestalde, horrelako hackathon ekimenei jarraipena ematea zaila bada ere, 2 urtean behin errepikatzeko proiektu polita litzateke. Izan ere, aukera paregabea da informatikaren eta hizkuntzalaritzaren arteko harremanak sendotzeko eta era praktikoa batean ikasteko.”



MANEX AGIRREZABALEkin

Literaturaren analisia eta sorkuntza jorratzen duen taldeko arduraduna da eta IXA taldeko doktorego ikaslea.

“Modu artistiko eta kurioso batean sozializatu nahi dugu hizkuntza naturalaren prozesamendua (HNP). Hasiera batean poesia aztertu nahi genuen baina txanogorritxuren ipuin klasikoarekin amaitu dugu jolasean. Umberto Ecoen Larrosaren izena istorioan agertzen den Salvatore pertsonaiaren ideian oinarrituta, aldeketan semantikoak egin dizkiegu istorioei, hala nola, hizkuntzak nahastuz, hitzak antonimoekin aldatuz etab. [Story buffet](#) izendatu dugu gure proiektua. Bestalde, hasierako helburua ez bazen ere, sorkuntza eta modifikazio horiek hizkuntza ikasteko baliagarriak izan daitezkeela ohartu gara.”

HELDU DA AZKEN EGUNA!

[codefest](#) • [uztaila 8, 2016](#)

Iritsi da CODEFEST 2016 udako eskolaren azken eguna eta horrekin batera, proiektu bakoitzaren aurkezpena entzuteko parada izan dugu:



Story buffet: Sortutako proiektua eta webgunea erakutsi digute antzerki-aurkezpen xelebretan.

Behagunea: Helburu nagusiak bete dituzte eta hizkuntza ezberdinak (katalana, italiarra) behagune proiektura gehitzeaz gain, sare sozialak eta geolokalizazio uztartu dituzte.

Morphological Analyser: Hainbat proiektu garatu dituzte aste honetan zehar, hungariarrentzat, buriaterarentzat (buriat, mongolieraren aldaera bat), Rifeko bereberrarentzat (Marokoko hizkuntza) eta uigurrera (Erdialdeko Asiako turkiar hizkuntza)

Quiz Bowl: Wikipedian oinarritutako QUIZera jolasteko aukera izan dugu: Ordenagailua vs. Gizakia. Oraingo honetan gizakien taldea gailendu da, baina estu-estu.

SRL, Semantic Role Labelling: Esaldi batean dauden sintaxi elementuak identifikatzeko garatu duten tresna aurkeztu digute. Proiektua etiketazio sistema batean oinarritu dute.

Machine Translation: Aurrerapen nabarmenak egin dituzte udako eskola honetan itzulpen automatikoaren inguruko hainbat proiektutan: frantsesa-okzitaniarra, tetuna-portugesara, euskara-gaztelera, ingelesa-galesa eta ingeles-euskara itzultzailea, besteak beste.

Zure TTS: Aste honen buruan lapurterako bertsioa martxan jartzeaz gain aurrerapen nabarmenak egin dute galikoarekin, batez ere, talde irlandarrekin egindako elkarlanari esker.



Aurkezpen interesgarri guztiak entzun ostean, talde bakoitzak beste proiektuak bozkatzeko aukera izan du. Quiz Bowl proiektua izan da txapelduna. Taldekide guztiak sari goxo bikaina jasotzeaz gain, Mans Huldenek eraman du txapela, taldeko arduradun gisa.



Amaitzeko taldeko argazki jendetsu bat atera dugu. Laster berriz ere batuko garelakoan...



PLAZER HUTSA IZAN DAI!