

Entzun Elhuyar aldizkaria

2014/04/01 **307. zenbakia** **ADIMEN ARTIFIZIALA**  
**PROGRAMAZIOA**

# Bota galdera!

**Lopez-Gazpio, Iñigo**

EHUko IXA Taldeko ikertzailea

Gogoan al duzue txikitan irakasleak gai bat amaitzean "ea, bota itzazue galderak!" esaten zuenean? Une hartan bertan hasten zen ezkutaketa-jolasa: aurreko ikaskidearen sorbalda atzean, mahai azpian, ikasgelako atearen atzealdean... edozein ihes-mugimendu baliozkoa



Arg. Andrei Kovalev (Follow)/123RF

zitzaigun irakaslearen begiradarekin topo ez egiteko. Lotsati horiei guztiei berri on bat dakarkiegu, gaurtik aurrera konputagailuak jasango baititu muxu-gorriak gure orde.

Ondorengo lerroetan, corpus bat oinarri hartuta esaldi-mailan galderak automatikoki eta masiboki sortuko dituen sistemaren hastapenak azalduko ditugu (sistema hauek QG (Question Generator) izenez ezagutzen dira). Horretarako, sukaldaritzan jaki goxo bat prestatzeko egiten dugun antzera, galdera-sorkuntzan ere beharrezkoa izango da pauso zehatz batzuei jarraitzea. Izan ere, galderak automatikoki sortzeko eta, adibidez, *pizza* bat egiteko jarraitu beharreko urratsak oso parekoak dira. Kontu hori guztia martxan jartzeko jarraituko dugun prozesuak lau urrats ditu, sekuentzialki jarraitu beharrekoak:

## 1. Sarrera-corpusa prozesatzea

Lehen urratsa, sarrera-corpusa prozesatzea --linguistikoki analizatzea-- helburu duena. Hortaz, bada, atal honetan, sarrera-corpusetik abiatuta ahalik eta **informazio gehiena** lortzen ahaleginduko gara. Informazio-bilaketa hori hizkuntzaren prozesamenduko hainbat eremutan

errepikatzen bada ere, funtsezko ataza da zentzuz aurrera jarraitu nahi badugu behintzat. Atal honi gogoz ekiteko, beharrezkoa izango dugu hizkuntza-baliabide desberdinak erabiltzea, hala nola analizatzaile sintaktikoak, analizatzaile morfologikoak, analizatzaile semantikoak, lematizatzaileak, tokenizatzaileak... Hala, jatorrizko esaldietan aditzak --jokatuak eta jokatu gabeak--, izenak, izenondoak, aditzondoak... identifikatzeko gai izango gara, eta, horrez gain, hitzak eta horiek osatzen dituzten izen-sintagmak, izen-sintagmen deklinabide-kasuak, aditzen arteko sintagma-konposaketak --aditz sinpleak, konposatuak, laguntzaileak... --eta esaldietako hitzen dependentzia sintaktikoak ere identifikatu beharko ditugu.

Sukaldaritzan aritzen garenean ere, konturatu gabe, hasieratze horren antzeko **analisi** bat egiten dugu: informazioa eta lanabesak prestatzen ditugu. Lehendabizi, jakia prestatzeko behar dugun eskuliburua hartzen dugu; ondoren, zer osagai behar ditugun irakurtzen dugu, haiek zerrendatu eta bilatzen ditugu; eta, gero, behar ditugun lanabesak sukaldeko mahaiaren gainean kokatzen ditugu, errezetaren pausoak trabarik gabe egiteko.

## 1. ADIBIDEA

Sarrera-corpora:

Pizzak asko gustatzen zaizkit; horregatik, afaltzeko hori prestatuko dut.

Analisi linguistikoaren emaitzak (bertsio murriztua):

<!-- Pizzak -->

<hitza id="t1" lemma="pizza" pos="N.IZE-ARR" case="ERG" />

<!-- asko -->

<hitza id="t2" lemma="asko" pos="A.ADB-ARR" />

<!-- gustatzen -->

<hitza id="t3" lemma="gustatu" pos="V.ADI-SIN" />

<!-- zaizkit -->

<hitza id="t4" lemma="izan" pos="V.ADL" />

<!-- horregatik -->

<hitza id="t5" lemma="horregatik" pos="A.ADB-ARR">

<!-- afaltzeko -->

<hitza id="t6" lemma="afaldu" pos="V.ADI-SIN" case="GEL">

<!-- hori -->

<hitza id="t7" lemma="hori" pos="D.DET-ERKARR" case="ABS">

<!-- prestatuko -->

<hitza id="t8" lemma="prestatu" pos="V.ADI-SIN">

<!-- dut -->

## 2. Galderen emaitzak hautatzea

Behin errezetako jakiak zerrendatu ditugunean --hobe gure hozkailuan egotea dendan baino!--, **haien bila joatea** da hurrengo urratsa. Tomatea eta urdaiazpikoa hozkailuan, gazta sukaldeko mahai gainean, ura iturrian, irina apalean, gatz armairuan... ; dena den, aitonaren baratzeke tomateak erabili **ordez** , aurreko eguneko tomate-saltsa erabiliko dugu, oso goxoa baitzegoen, eta, ziur, primeran geratuko dela.

Garatzen ari garen galdera-sortzailearen bigarren pauso honetan, sukaldean egin dugun modura, **galderen emaitza egokienen bila joatea da helburua** . Atal honetan ere, nahitaezkoa da hizkuntzaren prozesamendu naturaleko baliabideak erabiltzea, aurretik **sailkatu ditugun osagaien artean hoberenak hautatu** behar baititugu.

Hautaketa hori egiteko, informazio-iturri desberdinak erabili daitezke, esaterako: sare semantikoak; bizidunen eta bizigabeen hiztegiak edota artikuluen, albisteen, erakundeen, pertsonaia ospetsuen ... datu-baseak. Hala, hitz eta sintagma guztien artean, informazio gehien ematen digutenak aukeratuko ditugu, eta, ondorioz, sortuko ditugun galderak

interesgarriak izateko aukera gehiago izango dituzte.

## 2. ADIBIDEA

Behin-behineko hautagaien zerrenda (sukaldaritza-hiztegi sinple batean oinarrituz):

[ Pizzak ] → Hitzaren lema (pizza) semantika-hiztegian aurkitu dugu jaki gisa

[ asko ] → Informazio gehigarrik ez dugu lortu

[ gustatzen ] → Informazio gehigarrik ez dugu lortu

[ zaizkit ] → Informazio gehigarrik ez dugu lortu

[ horregatik ] → Informazio gehigarrik ez dugu lortu

[ afaltzeko ] → Hitzaren lema (afaldu) semantika-hiztegian aurkitu dugu jateko ekintza gisa

[ hori ] → Informazio gehigarrik ez dugu lortu

[ prestatuko ] → Informazio gehigarrik ez dugu lortu

[ dut ] → Informazio gehigarrik ez dugu lortu

Behin betiko hautagaien zerrenda:

[ Pizzak ] → Jakia (lema = pizza)

[ afaltzeko ] → Jateko ekintza (lema = afaldu)

## 3. Galdetzaileak aukeratzea

Gutxi falta da amaitzeko, labea berotzen hasi eta dagoeneko prestatzen ari garen jakia usaintzeko gai gara. Osagaiak mahai gainean ditugu,

beharko ditugun tresnak ere bai. Hortaz, eskuak zikintzeko garaia heldu zaigu, **osagai bakoitza prestatu** beharko baitugu: irin-orea osatu, gazta zuritu eta arraspatu, tomateari haziak kendu, urdaiazpikoari plastikozko banatzekoak kendu...

Orobat, galderak sortzeko jarraitzen ari garen prozesuan egin beharreko lanak antzekoak dira. Izan ere, corpora analizatu eta galderen emaitzak identifikatu ditugu; hortaz, aurrera jarraitzeko, eskuak zikindu beharko ditugu, birtualki bada ere, **sortuko ditugun galdera-motak prestatu**

behar baitira. Galdera-mota --eta horrekin batera, galdetzailea-- aukeratzea aurreko faseetan egindako lanarekin lotuta dagoen ataza da. Hala, sarrera-corpusean datak identifikatzea lortu badugu, zentzuzkoa litzateke denbora-motako galdera bat sortzea, "noiz" --edo tankera horretako-- galdetzaile bat erabiliz horretarako. Aitzitik, pertsonaia

### 3. ADIBIDEA

[ Pizzak ]

Galdera-mota → Bizigabe baten inguruko galdera, jakia alegia

Erabili beharreko galdetzailea → Zer

[ afaltzeko ]

Galdera-mota → Bizigabe baten inguruko galdera, jateko ekintza alegia

Erabili beharreko galdetzailea → Zer

ospetsuak identifikatzea lortu badugu, zentzuzkoa litzateke pertsonaia bati loturiko galdera bat osatzea, eta "nor" galdetzailea erabiltzea kasu horretan. Azaldu den gisa lan egingo dugu, aurreko fasean identifikatutako hautagai guztientzat galdera-mota zehaztu dugun arte.

#### 4. Galderak eraikitzea

Dena prest dugu, azken txanpa bakarrik geratzen da, hots, **egindako lan guztia bateratzea** ; hala eta guztiz ere, kontuz ibili behar dugu, azken uneko akats batek egindako lan guztia honda baitezake. Prestatu ditugun osagai guztiak erreparatu ondoren, lanari ekiten diogu: *pizza* ren orearen gainean tomatea zabaldu, gazta eta urdaiazpikoa gaintetik sakabanatu, eta, azkenik, jakia labean sartu. Orain, hogeit hamar minutu itxaron besterik ez dugu egin behar; bitartean, platerak, mahai-tresnak eta edalontziak prestatu baino ez dugu egin behar, labeko alarmak jakia ateratzeko ordua heldu dela ohartarazten digun arte.

Galdera-sortzailearen sistemaren azken urratsean ere, pareko lanak egin beharko ditugu. Hau da, jatorrizko esaldia hartu eta patroiz zein erregela batzuetan oinarrituz **galdera eraiki** behar baitugu. Horrez gain, beste



hainbat moldaketa ere egin beharko dira, esaterako, loturazko lokailuak ezabatu, puntuazio-markak egokitu, aditzen kasuak eta markak egokitu...; baina ez hori bakarrik, bururatzen zaizkigun beste hainbat gauza ere egin daitezke, galderak hobetzea helburu duen lan oro ongi etorria izango baita, hala nola anaforak ebaztea eta galderekin zerikusirik ez duen informazioa baztertzea --esaldiak laburtzea, alegia--.

#### 4. ADIBIDEA

Jatorrizko esaldia:

Pizzak asko gustatzen zaizkit; horregatik, afaltzeko hori prestatuko dut.

Jatorrizko esaldia orrazten:

- Esaldiak laburtu (puntuazio-ikurrak egokitu ditugu)
  1. Pizzak asko gustatzen zaizkit.
  2. Afaltzeko hori prestatuko dut.
- Anaforak ebatzi [ hori → pizza ]
  1. Pizzak asko gustatzen zaizkit.
  2. Afaltzeko pizza prestatuko dut.
- Galderak eraiki (aditzaren kasua ere egokituko dugu)
  1. Zer gustatzen zaio asko ? → [ Eraitza : Pizza ]
  2. Zer egingo du ? → [ Eraitza : Afaldu ]
  1. Zer prestatuko du afaltzeko ? → [ Eraitza : Pizza ]

## Ondorioak

Ikusi duzuenek, galderak masiboki sortzen dituen sistema martxan jartzeko, lau urrats egin behar dira: sarrera-corpusa prozesatu, galderen emaitzak hautatu, galdetzaileak aukeratu eta, azkenik, galderak eraiki.

*Pizza* bat egitea bezain erraza!

Garatu dugun sistema galdera-sortzailea --oso oinarrizko QG sistema dena-- eta ariketak automatikoki sortzeko gai diren beste hainbat sistema irakaskuntza-ikaskuntza arloarekin erabat loturik dauden baliabideak dira, irakasleek materiala prestatzeko behar duten denbora nabarmen murrizten baitute. Oro har, ariketak sortzeko sistema automatikoak e-learning plataforma handiagoetan integratuta egoten dira, eta, hala, lan-karga txiki baten ordainetan parte hartzen duten eta, aldi berean, ikasten dabiltzan erabiltzaile mordoak izateko aukera ematen dute.



Arg. <http://openclipart.org> webguneko irudietan oinarrituta sortua. Domeinu publikoan

Amaitzeko, hasieran aipatutako lotsati guzti horiei, jada badakizue zuen lotsak pizza bat egitea bezain erraz desager daitezkeela (hogei minutu baino denbora gehixeago beharko baduzue ere); beraz, atera zaitezte

zuen ezkutalekuetatik, eta **bota itzazue galderak !**

## Bibliografia

Wikipediako webgune ofiziala, <http://www.wikipedia.org>.

IXA taldeko Eihera aplikazioaren gunea, <http://ixa.si.ehu.es/Ixa/Produktuak/1273220198>

IXA taldeko Eulia aplikazioaren gunea, <http://ixa.si.ehu.es/Ixa/Produktuak/1274694158>

IXA taldeko Ixati aplikazioaren gunea, <http://ixa.si.ehu.es/Ixa/Produktuak/1273220525>

Aldabe, I.; Lopez de Lacalle, M.; Maritxalar, M.; Martinez, E.; Uria, L.: ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques. UPV/EHU.

Yao, X.; Bouma, G.; Zhang, Y.: Semantics-based Question Generation and Implementation. Johns Hopkins University, University of Groningen. Saarland University.

## Idatzi zuk zeuk Gai librean atalean

Gai librean aritzeko, bidali zure artikulua [aldizkaria@elhuyar.com](mailto:aldizkaria@elhuyar.com) helbidera

Hauek dira Gai librean atalean Idazteko arauak

