

AnHitz

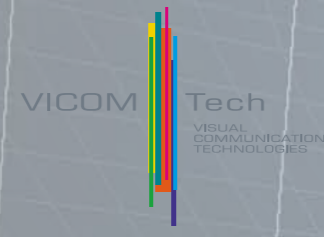
Euskararentzako hizkuntza-, ahots-, eta ikusmen-teknologiaren garapena eta integrazioa

AnHitz euskararentzako hizkuntza- eta ahots-teknologiak garatzeko Eusko Jaurlaritzako Industria sailak lagundutako proiektua da.

AnHitz-eko partehartzaileak oso profil ezberdinetako ikerketa taldeak dira: testu-prozesamendua, ahotsaren prozesamendua eta multimedia.

Kide hauek osatzen dute AnHitz: VICOMTech eta Robotiker ikerketa-zentroek, Elhuyar Fundazioko I+G taldeak eta EHUKO IXA eta Aholab ikerketa-taldeek.

Proiektuaren orain arteko emaitzak 7 hizkuntza baliabide, lengoia naturalaren prozesamendurako 13 tresna eta 7 aplikazio dira.



AnHitz proiektuan garatutako teknologiak

Tresnak

Erauzterm - Euskarazko testu eta corpusetako terminologia erauzketa automatikoa.

ElexBI - Gaztelania-Euskara itzulpen-memoretatik termino pare baliokideen erauzketa.

Corpusgile eta Eulia - Corpusak sortu, etiketatu eta kontsultatzeko tresna aurreratuak.

CorpEus - Euskararentzako web-as-corpus tresna.

Dokusare - Ingurune eleaniztun baten eduki antzekoa duten zientzia-berriak identifikatzeko tresna.

Co3 - Corpus konparagarri eleaniztunak Internetetik automatikoki sortzeko tresna.

AzerHitz - Gaztelania-Euskara corpus konparagarrietatik termino pare baliokideen erauzketa.

Elezkari - Euskara, gaztelania eta ingelesarentzako informazio-bilaketa eleanitza.

Eulibeltz - Corpus lerrokatu elebidunen sorkuntza eta etiketatzea.

AhoT2P - Euskara estandarerako letra-alofono transkribatzailea.

AhoTTS_Mod1 - Ahotsaren sintesirako prozesagailu linguistikoa.

Vocon3200 - Gramatikan oinarritutako ahots-egazutza sistema euskararentzat.

Pertsonaia birtualak - Emozioak eta ahotsa duten 3D pertsonaiak.

Baliabideak

ZT Corpora - 8,5 milioi hitzeko euskarazko testu espezializatuen bilduma.

EPEC - Etiketatu eta desanbiguatutako 300.000 hitzeko corpora.

SpeechDat FDB1060-EU - Telefono finkotik grabatutako euskarazko 1.060 grabazio.

SpeechDat MDB600-EU - Telefono mugikorretik grabatutako euskarazko 660 grabazio.

EMODB - Emoziodun ahots datu-basea, emakumezko ahotsarekin.

Amaia and Aitor - Emoziodun ahots datu-basea.

BIZKAIFON - Mendebaldeko euskararen datu-base multimodala (ahotsa eta bideoa).

Aplikazioak

Xuxen - Euskarazko zuzentzaile ortografikoa.

Lemantizazio bidezko hiztegiak.

Elebila - Interneten euskarazko edukia aurkitzeko bilatzaile publikoa.

Opentrad-Matxin - Iturburu-kode irekiko Gaztelania-Euskararako itzulpen automatikoko sistema.

Ingelesa-Euskara itzulpen automatikoa - Itzulpen automatiko estatistikoa.

AhoTTS - Testu-ahots bihurtzaile modularra euskara eta gaztelaniarentzat.

AhoTTS PDarentzat - PDA gailuentzako aplikazio multiplataforma.

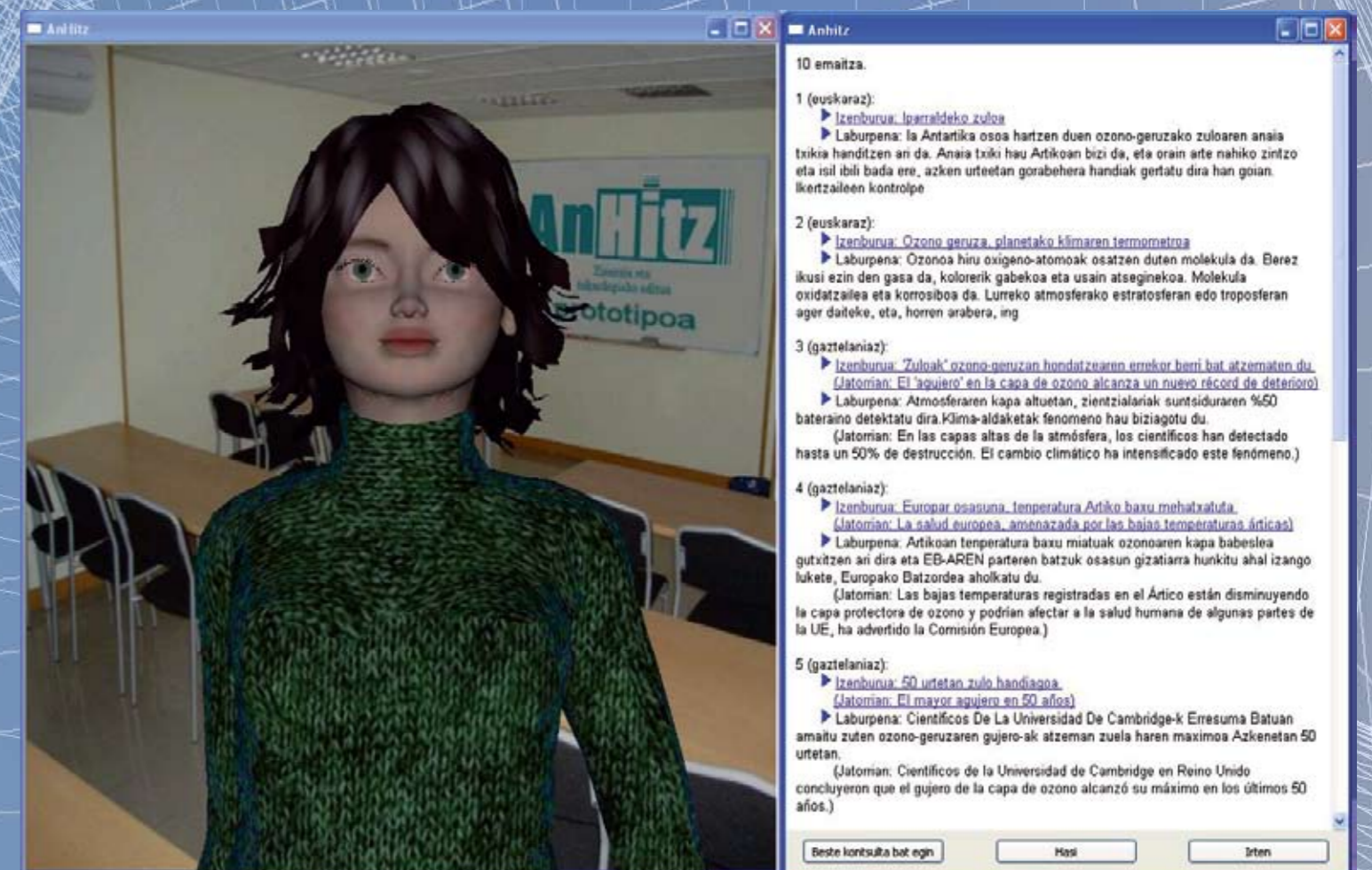
Demostrazio eszenatokia

Teknologiaren integrazioa

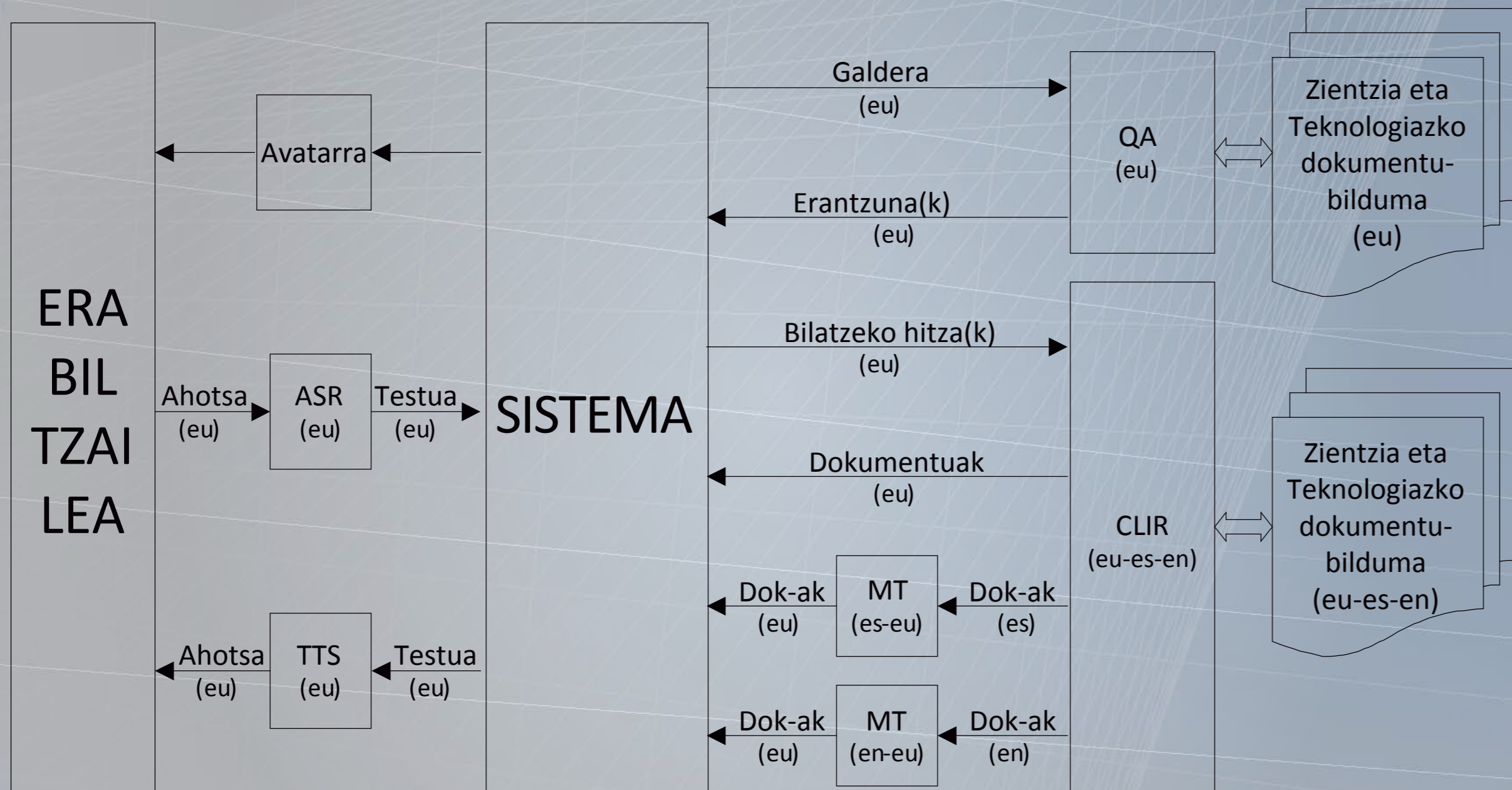
Proiektuaren azken urtean, lehenbiziko aldiz integratu ditugu baliabide eta tresna hauek, lengoia natural bidezko komunikazio-interfazea duen eduki-kudeaketarako aplikazio eleaniztun bat sortuz.

Aplikazio hau Zientzia eta Teknologia gaietako Galdera-Erantzute eta Informazio-Bilaketa Eleaniztun sistema bat da.

Sistemaren eta erabilzailearen arteko interakzioa euskaraz da (euskarazkoak ez diren informazio-bilaketaren emaitzak itzulpen automatiko bidez itzultzen dira), ahotsaren sintesia, ahotsaren egazutza eta interfaze bisual bat erabiliz.



Sistemaren arkitektura



Teknologiak

Question Answering
(galderak erantzutea)

Cross Lingual Information Retrieval
(informazio-bilaketa eleaniztuna)

Machine Translation
(itzulpen automatikoa)

Speech Synthesis
(ahotsaren sintesia)

Automatic speech Recognition
(ahotsaren egazutza)

Visual Interface
(interfaze bisualak)