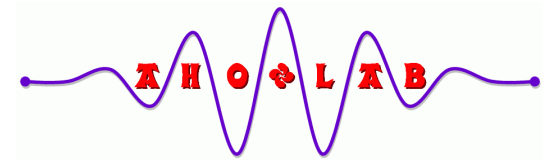


Web Communication Protocols for Coordinating the Modules of AnHitz, a Basque-Speaking Virtual 3D Expert on Science and Technology



Igor Leturia – Elhuyar Foundation
WSPP2010 workshop – 2010/07/17-18 – Valetta, Malta

Outline

- Background
- The AnHitz demo
- Communication model
- Conclusions

Outline

- **Background**
- The AnHitz demo
- Communication model
- Conclusions

The AnHitz project

- A strategic research project
- On language, speech and visual technologies
- Mainly for Basque language
- Partly funded by the Industry and Culture departments of the Basque Government
- Carried out by a consortium of five members
- In the years 2006-2008

The AnHitz consortium (I)

Background

The AnHitz demo

Communication model

Conclusions

- VICOMTech (IK4)
 - Research center working in interactive computer graphics and digital multimedia
- Robotiker (Tecnalia)
 - Technology center specialized in information and telecommunication technologies

The AnHitz consortium (II)

- Elhuyar Foundation
 - A nonprofit making organization for promotion of Basque language in science and technology (dictionaries, corpora, R&D on language technologies...)
- IXA Group (Basque Country University)
 - Research group specialized on Natural Language Processing
- Aholab Group (Basque Country University)
 - Research group specialized on Speech Technologies

Results of AnHitz (I)

- Textual resources
 - Monolingual corpora, both general and specialized, with different annotation levels
- Speech resources
 - Speech databases of various speakers, emotions, registers and recording conditions

Results of AnHitz (II)

Background

The AnHitz demo

Communication model

Conclusions

- Textual tools
 - Terminology extraction from monolingual, parallel and comparable corpora
 - Corpus tools
 - Web as corpus tools
 - Information retrieval (monolingual and cross-lingual)
 - Entity recognition
 - Machine translation
 - Question answering
 - Spell checking

Results of AnHitz (III)

Background

The AnHitz demo

Communication model

Conclusions

- Speech tools
 - Text to speech tools
 - Automatic speech recognition
 - Speaker recognition
- Visual tools
 - Avatar with spoken communication

Outline

- Background
- The AnHitz demo
- Communication model
- Conclusions

Outline

- Background
- **The AnHitz demo**
- Communication model
- Conclusions

The AnHitz demo

- Integration of various language, speech and visual technologies developed in the AnHitz project
- Interaction via speech in Basque, with a 3D virtual character
- Expert on science and technology
 - Answer questions
 - Performs cross-lingual information retrieval and translates the results to Basque

Technology integration (I)

Background

The AnHitz demo

Communication model

Conclusions

- 3D avatar showing emotions (VICOMTech)
- Multilingual and emotional text-to-speech synthesizer (Aholab)
- Basque speech recognition (Robotiker)
- Grammars for recognition of scientific questions and terms in Basque (Elhuyar)
- Multilingual science and technology corpora (Elhuyar)

Technology integration (II)

Background

The AnHitz demo

Communication model

Conclusions

- Basque question answering (IXA)
- Cross-lingual information retrieval for Basque terms (Elhuyar)
- Spanish-Basque and English-Basque Machine translation (IXA)
- Main program to guide the flow of the conversation (Elhuyar)





AnHitz

10 emaitza.

1 (euskaraz):

- ▶ [Izenburua: Iparraldeko zuloa](#)
- ▶ Laburpena: la Antartika osoa hartzen duen ozono-geruzako zuloaren anaia txikia handitzen ari da. Anaia txiki hau Artikoan bizi da, eta orain arte nahiko zintzo eta isil ibili bada ere, azken urteetan gorabehera handiak gertatu dira han goian. Ikertzaileen kontrolpe

2 (euskaraz):

- ▶ [Izenburua: Ozono geruza, planetako klimaren termometroa](#)
- ▶ Laburpena: Ozonoa hiru oxigeno-atomoak osatzen duten molekula da. Berez ikusi ezin den gasa da, kolorerik gabekoa eta usain atseginekoa. Molekula oxidatzailea eta korrosiboa da. Lurreko atmosferako estratosferan edo troposferan ager daiteke, eta, horren arabera, ing

3 (gaztelaniaz):

- ▶ [Izenburua: 'Zuloak' ozono-geruzan hondatzearen errekor berri bat atzematzen du](#)
([Jatorrian: El 'agujero' en la capa de ozono alcanza un nuevo récord de deterioro](#))
- ▶ Laburpena: Atmosferaren kapa altuetan, zientzialariak suntsiduraren %50 bateraino detektatu dira. Klima-aldaketak fenomeno hau biziagotu du.
(Jatorrian: En las capas altas de la atmósfera, los científicos han detectado hasta un 50% de destrucción. El cambio climático ha intensificado este fenómeno.)

4 (gaztelaniaz):

- ▶ [Izenburua: Europar osasuna, tenperatura Artiko baxu mehatxatuta](#)
([Jatorrian: La salud europea, amenazada por las bajas temperaturas árticas](#))
- ▶ Laburpena: Artikoan tenperatura baxu miatuak ozonoaren kapa babeslea gutxitzen ari dira eta EB-AREN parteren batzuk osasun gizatiarra hunkitu ahal izango lukete, Europako Batzordea aholkatu du.
(Jatorrian: Las bajas temperaturas registradas en el Ártico están disminuyendo la capa protectora de ozono y podrían afectar a la salud humana de algunas partes de la UE, ha advertido la Comisión Europea.)

5 (gaztelaniaz):

- ▶ [Izenburua: 50 urtetan zulo handiagoa](#)
([Jatorrian: El mayor agujero en 50 años](#))
- ▶ Laburpena: Científicos De La Universidad De Cambridge-k Erresuma Batuan amaitu zuten ozono-geruzaren gujero-ak atzeman zuela haren maximoa Azkenetan 50 urtetan.
(Jatorrian: Científicos de la Universidad de Cambridge en Reino Unido concluyeron que el gujero de la capa de ozono alcanzó su máximo en los últimos 50 años.)

Beste kontsulta bat egin Hasi Irten

Outline

- Background
- The AnHitz demo
- Communication model
- Conclusions

Outline

- Background
- The AnHitz demo
- **Communication model**
- Conclusions

Heterogeneous modules

Background

The AnHitz demo

Communication model

Conclusions

- Very different modules
 - Built by different organizations
 - Run on different operating systems
 - Programmed in different programming languages
 - Use different libraries

Web communication

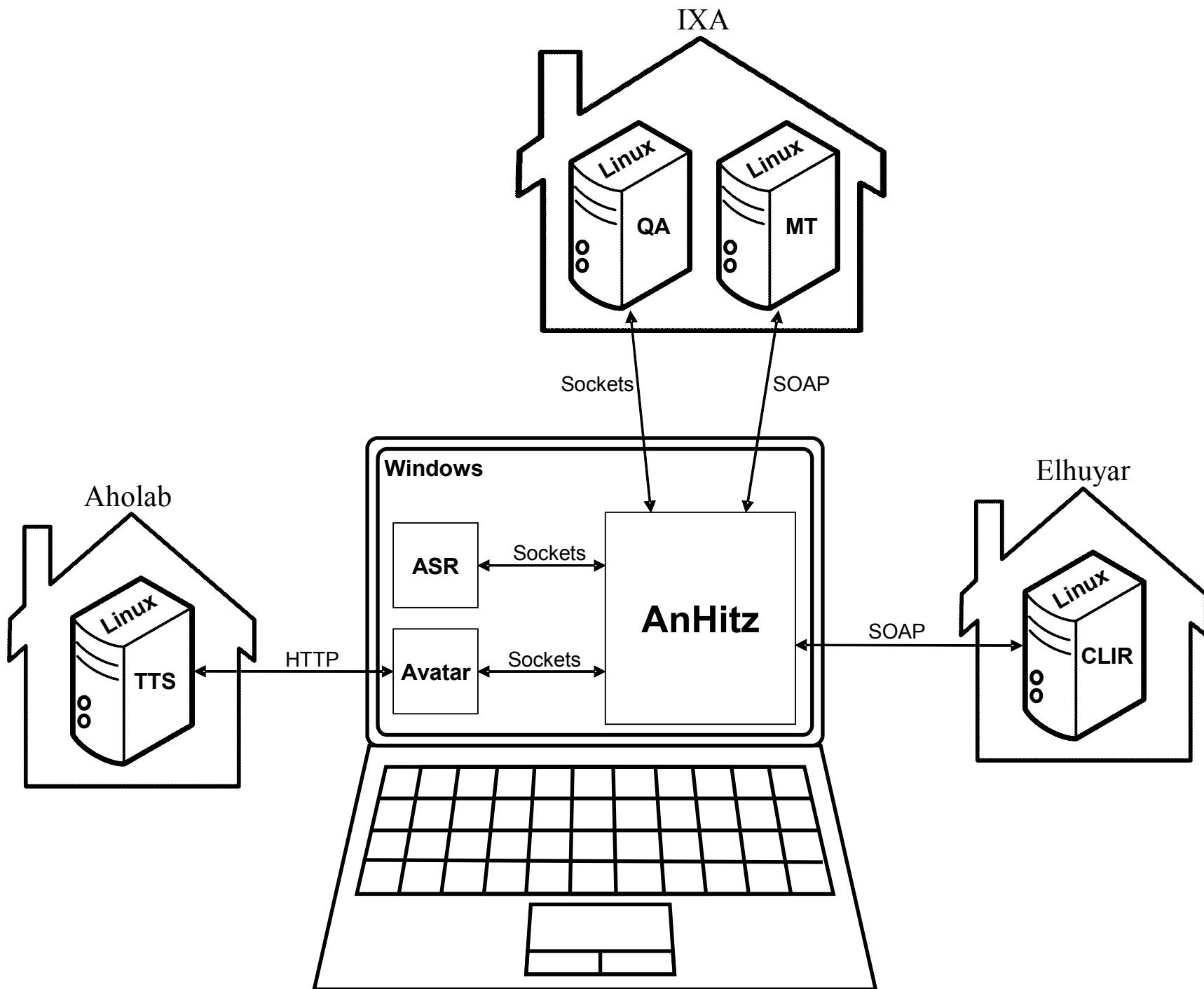
Background

The AnHitz demo

Communication model

Conclusions

- Distributed system with modules running in different machines over the Internet
- Communication via net protocols and services, even among the modules running in the same machine



Outline

- Background
- The AnHitz demo
- Communication model
- Conclusions

Outline

- Background
- The AnHitz demo
- Communication model
- **Conclusions**

Non technical

- Great interest from media and public
 - Press conference on every radio, TV or newspaper
 - Many media dedicated a video, interview or article
 - Students and general public played with it in stands in the Week of Science and Technology
- Good evaluation results
- Collaboration between agents working in different areas is necessary to exploit the whole potential of language technologies

Positive aspects

- Enormous benefits of the modular architecture and web communications approach for integrating such different technologies
- Doubt if it could have been done otherwise
- Modules easily replaceable to their newest versions

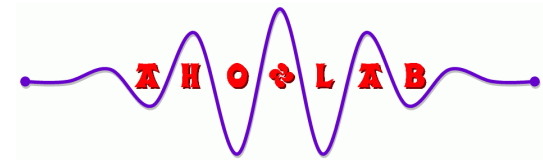
Negative aspects

- Internet connection needed
- Affected by connection speed
- Latency problems, mainly for having to download the audio files of the TTS over the Internet
 - Solved to some extent by locally caching most repetitive and already said sentences
- Lack of control over the remote modules
 - When something goes wrong in one, the whole demo is affected and it is not trivial to put things back up

Future work

- Possible solution: virtualization
 - Not sure if possible (complexity of libraries...)
 - Still network communication would be used
- Put the demo in the web
 - Difficulties: voice recognition in web browser...
- End-user application
 - Tourist information offices, science museums...
 - Long way for making the demo an end-user application

Web Communication Protocols for Coordinating the Modules of AnHitz, a Basque-Speaking Virtual 3D Expert on Science and Technology



Igor Leturia – Elhuyar Foundation
WSPP2010 workshop – 2010/07/17-18 – Valetta, Malta