



**Sistemas Inteligentes de
Acceso a la Información**

SINAI (Intelligent Systems for Information Access) research group of the University of Jaén (Spain) is a team composed by 10 researchers and two PhD students. SINAI is specialized in the study of Human Language Technologies (HLT) also known as Computational Linguistics. Since 2000, the intensive work of the SINAI team has allowed the opening of different lines of research. Some of the most important are:

- Multilingual and Multimodal Information Retrieval
- Document categorization
- Data and text mining
- Information extraction
- Geographical information retrieval
- Opinion mining

The acquired knowledge by SINAI is shown in all the research papers published, the research resources developed, the research projects carried out, and the different industrial collaborations.

What can SINAI do?

The best way to know the capabilities of SINAI is knowing the projects carried out:

R2D2: Retrieval of answers in Digital Documents

- **Goal:** The development of a multilingual information retrieval system of digital documents
- **Skills applied:**
 - Multilingual Information retrieval
 - Question Answering
 - Document management
- **Scope:** National (Spain)

Text-Mess: Intelligent, Interactive and Multilingual Text Mining based on Human Language Technologies



- **Goal:** Developing of intelligent, interactive and multilingual techniques for text mining tasks. These new techniques were applied in search engines with the purpose of discovering the need behind the user query.
- **Skills applied:**
 - Information retrieval
 - Information extraction
 - Multimodal (images, audio, video) information retrieval.

- **Scope:** National (Spain)

Text-Mess 2.0: The New Digital Media Challenges Facing Human Language Technologies

The logo for TextMESS2.0 features the text 'TextMESS2.0' in a blue, sans-serif font. The '2.0' is notably larger and more prominent than the rest of the text. Below the main text, there is a faint, semi-transparent reflection of the same text.

- **Goal:** Development of methods for identifying subjective information with the purpose of adding this new methods to search engines.
- **Skills applied:**
 - Opinion Mining → Subjective classification
 - Information Retrieval
 - Information extraction
- **Scope:** National (Spain)

GeOasis: Turistic information Intelligent System based on geographical position

The logo for GeOasis features the word 'GeOasis' in a stylized, green, cursive font. The 'G' is larger and more prominent, and the 'O' is a circle that overlaps with the 'a'. The text has a slight shadow effect.

- **Goal:** Development of an interactive system for turistic information retrieval system based on the current geographical position of the user.
- **Skills applied:**
 - Information retrieval
 - Entity recognition
 - Speech recognition
 - Geographical information retrieval
- **Scope:** Regional (Andalusia)

AORESCU: Opinion Mining in Social Networks

- **Goal:** Development of an opinion mining system at entity level taking as sources the information from the Internet and the Web 2.0.
- **Skills applied:**
 - Spanish Opinion mining → Subjectivity and polarity classification
 - Entity extraction
 - Information extraction
 - Information monitoring
 - Domain adaption techniques
- **Scope:** Regional (Andalusia)

ATTOS: Trends & Subjects Analysis Through Opinions & Feelings



- **Goal:** Development of a web 2.0 monitor system which is capable of generating reports about opinions trends related to a topic or entity.
- **Skills applied:**
 - Opinion mining → Subjectivity and polarity classification
 - Entity recognition
 - Multilingual information retrieval
 - Geographical information retrieval
 - Information monitoring
 - Information generation → text summarization, opinion summarization
 - Trends analysis
 - Temporal series analysis
- **Scope:** National (Spain)

FIRST: A Flexible Interactive Reading Support Tool



- **Goal:** Development of an intelligent system to facilitate the reading comprehension to people with autism.
- **Skills applied:**
 - Multimodal information retrieval
 - Figurative language treatment
 - Text summarization
 - Anaphora resolution
 - Text simplification
 - Word sense disambiguation
- **Scope:** International (FP7)

Developed resources

The team of SINAI has developed several resources, some of the most relevance are:

- Opinion mining:
 - iSOL → Spanish opinion bearing words list.
 - MCE Corpus → Opinion mining corpus of English movies reviews. This is a parallel corpus of the Spanish corpus of movies reviews MuchoCine.
 - SINAISaCorpus → Opinion mining corpus in Spanish of product reviews from Amazon.
 - OCA Corpus → Opinion mining corpus of Arabic movie reviews

- [EVOCA Corpus](#) → Opinion mining corpus of English movie reviews.
This is a parallel corpus of the Arabic corpus of movies reviews OCA
- Text classification
 - [TeCat](#) → Tool for automatic text multi-tag classification
 - [HEP Corpus](#) → Corpus for multi-tag text classification
- Question answering
 - [Spanish QC](#) → Spanish Corpus for question answering systems.

Collaborations

Academic

Spain

University of Alicante
 Polytechnic University of Valencia
 Polytechnic University of Barcelona
 University of Barcelona
 University Carlos III of Madrid
 University of Sevilla
 University of Huelva
 European University of Madrid
 University of Basque Country
 Cervantes Institute
 University of Distance Education

International

University of Wolverhampton (UK)
 Center for Applied Linguistics (Cuba)
 Joint Research Center Language
 Technology (Italy)
 INESC-ID Engineering Institute of
 Computational Systems, Research and
 Development (Portugal)
 INAOE – National Institute of Astrophysics
 and Optics (Mexico)

Industrial

- [GuadalTel](#)
- [Yahoo! Research](#)
- [Vocali](#)
- [Grammata](#)
- [Daedalus](#)
- [Novasoft](#)
- [IWEB](#)
- [Jaén city council](#)
- [Jaen province government](#)