

25th edition of the Annual Conference of the Spanish Society for Natural Language Processing 2009 (SEPLN'09)

Monday, 7th of September:

- Workshop 1: Web as a Corpus 2009 (WAC5, <http://www.sigwac.org.uk/wiki/WAC5>)
- Workshop 2: Information Retrieval and Information Extraction for Less Resourced Languages IE-IR-LRL (SALTMIL <http://ixa2.si.ehu.es/saltmil/en/activities/lrec2008/sepln-2009-workshop-cfp.html>)

Thursday, 10th of September, afternoon:

- Workshop 3: Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 2009).

1st day (Tuesday, 8th of September)

09:00-9:30	Registration	
9:30-10:00	Opening	
10:00-11:00	Invited talk: Piek Vossen The KYOTO project: an implementation of the Global Wordnet Grid	
11:00-11:20	Coffee break	
11:20-13:00	Session 1: Development of linguistic resources and tools <ul style="list-style-type: none"> • Inducción de un Lexicón de Opinión Orientado al Dominio. Fermín Cruz, José A. Troyano, Javier Ortega, Carlos G. Vallejo • Detección de Expresiones Temporales TimeML en Catalán mediante Roles Semánticos y Redes Semánticas. Héctor Llorens, Borja Navarro, Estela Saquete • Hacia un sistema de clasificación automática de sustantivos deverbales. Horacio Rodríguez, Mariona Taulé, Aina Peris • Geo-NER: un reconocedor de entidades geográficas para inglés basado en GeoNames y Wikipedia. José Manuel Perea Ortega, Arturo Montejo Ráez, Fernando Martínez Santiago y L. Alfonso Ureña López 	Session 2: Speech processing <ul style="list-style-type: none"> • Managing Speaker Identity and User Profiles in a Spoken Dialogue System. J.M. Lucas, F. Fernández, J. Salazar, J. Ferreiros and R. San Segundo • Algoritmo de inserción de pausas para una lengua declinada. Eva Navas, Iñaki Sainz, Jon Sanchez, Ibon Saratxaga, Inmaculada Hernáez • On the definition of a prosodically balanced corpus: combining greedy algorithms with expert guided manipulation. David Escudero, Lourdes Aguilar, Antonio Bonafonte, Juan Maria Garrido • Tratamiento del léxico del euskara occidental basado en la división de radical y desinencia para reconocimiento de habla dialectal. Igor Odriozola, Eva Navas, Jon Sánchez, Inma Hernaez
13:00-15:00	Lunch	
15:00-16:40	Session 3: Information retrieval and extraction <ul style="list-style-type: none"> • Propuesta y evaluación de un método de generación de resúmenes extractiva basado en conceptos en el ámbito biomédico. Manuel de la Villa y Manuel J. Maña • On the Relevance of Search Space Reduction in Automatic Plagiarism Detection. Alberto Barrón Cedeño, Paolo Rosso • Reconocimiento de definiciones asociadas a frases predicativas en contextos definitorios. César Aguilar, Gerardo Sierra • Describing Biomedical Document Sets in Terms of its Most Distinctive Facts. Yunior Ramírez-Cruz, Rafael Berlanga-Llavori, Aurora Pons-Porrata 	Session 4: Machine translation <ul style="list-style-type: none"> • Development of a Basque to Spanish machine translation system. Mireia Ginestí-Rosell, Gema Ramírez-Sánchez, Sergio Ortiz-Rojas, Francis M. Tyers, Mikel L. Forcada • Evaluación de un sistema de traducción automática basado en reglas o por qué BLEU sólo sirve para lo que sirve. Aingeru Mayor, Iñaki Alegria, Arantza Díaz de Ilarraza, Gorka Labaka, Mikel Lersundi, Kepa Sarasola • The HiFST System for the EuroParl Spanish-to-English Task. Gonzalo Iglesias, Adrià de Gispert, Eduardo R. Banga, William Byrne • Procesamiento Lingüístico en Métricas de Evaluación Automática de Traducciones. Enrique Amigó, Jesús Giménez y Felisa Verdejo
16:40-17:00	Coffee break	
17:00-18:30	Projects: <ul style="list-style-type: none"> • RILG: Recursos Integrados da Lingua Galega. Xavier Gómez Guinovart, Antón Santamarina • KNOW: Developing large-scale multilingual technologies for language understanding. Eneko Agirre, German Rigau, Irene Castellón, Laura Alonso, Salvador Climent, Lluís Padro, Montse Cuadros, Salvador Climent, Marta Coll-Florit • Carvalho: English-Galician SMT system from EuroParl English-Portuguese parallel corpus. José Ramom Pichel, Paulo Malvar, Oscar Senra, Pablo Gamallo, Alberto García • Fostering Language Resources Network: FLReNet. Núria Bel and Nicoletta Calzolari • QEAVIS: Quantitative Evaluation of Academic Websites Visibility. M. F. Verdejo, E. Amigó, L. Araujo, V. Fresno, G. Garrido, R. Martínez, A. Peñas, A. Pérez, J.R. Pérez, A. Rodrigo, J. Romo, A. Zubiaga, I. Aguillo, M. Fernández, A. M. Utrilla • Conclusiones de la primera Jornada del Processament Computacional del Català. Gemma Boleda, Montse Cuadros, Cristina España-Bonet, Maite Melero, Lluís Padró, Martí Quixal, Carlos Rodríguez • Kyoto Project. Eneko Agirre, Arantza Casillas, Arantza Díaz de Ilarraza, Ainara Estarrona, Kike Fernández, Koldo Gojenola, Egoitz Laparra, German Rigau, Aitor Soroa • Producción eficiente de recursos lingüísticos: el proyecto Victoria. Lionel Nicolas, Miguel A. Molinero, Benoît Sagot, Elena Sánchez Trigo, Eric de la Clergerie, Miguel Alonso Pardo, Jacques Farré, Joan Miquel Verges 	

18:30-19:15 cocktail

19:15-20:15 Music concert: Esfumato (Palacio Miramar)

2nd day (Wednesday, 9th of September)

08:45-9:00	Registration
9:00-10:40	<p>Session 5: Automatic analysis of textual contents</p> <ul style="list-style-type: none"> • Clasificación de Páginas Web con Anotaciones Sociales. Arkaitz Zubiaga, Raquel Martínez, Víctor Fresno • Características y rasgos afectivos del humor: Un estudio de reconocimiento automático del humor en textos escolares en catalán. Antonio Reyes, Paolo Rosso, Antònia Martí, Mariona Taulé • Fuzzy clusterers combination by positional voting for robust document clustering. Xavier Sevillano, Joan Claudi Socoró, Francesc Alías • Caracterización de una entidad basada en opiniones: un caso de estudio. Damiano Spina, Enrique Amigó, Julio Gonzalo, Bernardino Beotas
10:40-11:00	Coffee break
11:00-12:00	<p>Invited talk: Horacio Rodríguez Conocimiento rico, conocimiento pobre y desconocimiento en PLN. El ejemplo de la implicación textual.</p>
12:10-13:45	<p>Demonstrations:</p> <ul style="list-style-type: none"> • Zirano Cúpula, una aplicación de las variables conceptuales según la semántica conceptual multilingüaje. Miguel Santolaria García, Ana Santolaria Gracia, Carmen Santolaria Gracia, Marco Calvo Lafarga, José Ángel García Landa • Construcción de un sistema PLN usando el framework UIMA. Fermín Cruz, José A. Troyano, Fernando Enriquez, Víctor J. Díaz • INVENIO: búsqueda semántica basada en tecnología Google. Pedro Vivancos Vicente, Juan Salvador Castejón Garrido, David García Gil • La herramienta IDRA (Indexing and Retrieving Automatically). Rubén Granados Muñoz, Ana García Serrano, José M. Goñi Menoyo • Baratinoo speaks Spanish as well. Elixabete Murguia, Thierry Moudenc, Paul Bagshaw • Morfología de estados finitos en software libre: aplicación al euskera. Iñaki Alegria, Izaskun Etxeberria, Nerea Ezeiza, Montse Maritxalar • AnHitz, integración de tecnologías de la lengua dentro de un prototipo de experto virtual en ciencia y tecnología. Kutz Arrieta, Arantza Diaz de Ilarraza, Inma Hernández, Urtza Iturraspe, Igor Leturia, Eva Navas, Arantza del Pozo, Kepa Sarasola • Bitextor, un cosechador automático de memorias de traducción a partir de sitios web multilingües. Miquel Esplà Gomis • Co-Co, a web interface for corpora compilation. Cristina España-Bonet, Marta Vila, Horacio Rodríguez and M^a Antònia Martí • OAL: una herramienta para el desarrollo de recursos lingüísticos. Javier Couto, Helena Blancafort, Somara Seng, Anass Talby, Claude de Loupy • Una aproximación para la adquisición de un corpus de diálogo. L.F. Hurtado, E. Segarra, E. Sanchis, F. García, D. Griol
13:45-15:15	Lunch
15:15-16:30	<p>Panel: XXV years of SEPLN: past, present, and strategies for the future:</p> <ul style="list-style-type: none"> • Felisa Verdejo (UNED, Madrid) • Toni Martí (CLiC Centro de Lenguaje y Computación; Universidad de Barcelona) • Manuel Palomar (Universidad de Alicante) • Luis de Sopeña (IBM, Madrid) • Alfonso Ureña (Universidad de Jaén) • Arantza Díaz de Ilarraza (Universidad del País Vasco)
16:30-16:45	Coffee break
16:45-17:45	SEPLN meeting (SEPLN members)

21:00 Conference dinner

3rd day (Thursday, 10th of September)

09:30-11:00	<p>Papers (poster presentation):</p> <ol style="list-style-type: none"> 14. Una gramática de dependencias basada en patrones de etiquetas. <i>Pablo Gamallo Otero, Isaac González</i> 15. Determiner errors in Basque: Analysis and Automatic Detection. <i>Larraitx Uria, Bertol Arrieta, Arantza Díaz de Ilarraza, Montse Maritxalar, Maite Oronoz</i> 16. Bringing Role and Reference Grammar to natural language understanding. <i>Carlos Perriñán Pascual y Ricardo Mairal Usón</i> 17. Comparación de tres Modelos de Representación de Texto en la Generación Automática de Resúmenes. <i>Romyña Montiel Soto, René Arnulfo García Hernández, Yulia Ledeneva, Rafael Cruz Reyes</i> 18. Extracción crosslingüe de documentos usando mapas semánticos no-lineales. <i>Rafael E. Banchs, Marta R. Costa-jussà</i> 19. Anotación Pragmática de Diálogos en XML. <i>Dolores Cuadra, Mario Crespo, Javier Calle</i> 20. Towards a Rich Dependency Annotation of Spanish Corpora. <i>S. Mille, A. Burga, V. Vidal, L. Wanner</i> 21. Aplicación de la realimentación por relevancia con base textual para sistemas de recuperación visual. <i>Arturo Montejo Ráez, José Manuel Perea Ortega, Manuel Carlos Díaz Galiano, L. Alfonso Ureña López</i> 22. Detección de vocales mediante modelado de clusters de fonemas. <i>Iker Luengo, Eva Navas, Jon Sánchez, Inma Hernández</i> 23. Fast approximate string matching with finite automata. <i>Mans Hulden</i> 24. Construcción y extensión de un léxico morfológico y sintáctico para el Español: el Leffe. <i>Miguel A. Molinero, Benoît Sagot, Lionel Nicolas</i> 25. Mejora de la Interfaz de Entrada de Sistemas de Diálogo Usando Modelos Contextuales y Reglas Gramaticales. <i>Ramón López-Cózar, Zoraida Callejas</i> 26. Bases para evaluar la anotación de corpus de emociones espontáneas. <i>Zoraida Callejas, Ramón López-Cózar</i>
11:00-12:00	<p>Invited talk: Joakim Nivre Current Trends in Data-Driven Dependency Parsing</p>
12:00-12:20	<p>Coffee break</p>
12:30-13:45	<p>Session 6: NLP and Evaluation</p> <ul style="list-style-type: none"> • Comparación de Enfoques para Evaluar la Validación de Respuestas. <i>Álvaro Rodrigo, Anselmo Peñas, Felisa Verdejo</i> • Evaluación del algoritmo de fusión documental 2stepRSV en un sistema de Búsqueda de Respuestas Multilingüe. <i>Miguel Angel García Cumberas, Fernando Martínez Santiago, L. Alfonso Ureña López y José Manuel Perea Ortega</i> • Evaluación de métodos semi-automáticos para la conexión entre FrameNet y SenSem. <i>Laura Alonso, Irene Castellón, Egoitz Laparra, German Rigau</i>
13:45-14:00	<p>Closing</p>

Invited talks

Piek Vossen, VU University Amsterdam. Tuesday, 8th of September. 10:00-11:00

The KYOTO project: an implementation of the Global Wordnet Grid

In this presentation, I will describe the KYOTO project, which is funded by the European Union and by national funds in Taiwan and Japan. The goal of KYOTO is a system that allows people in communities to define the meaning of their words and terms in a shared Wiki platform so that they become anchored across languages and cultures but also so that a computer can use this knowledge to detect knowledge and facts in text. The languages of the project are English, Dutch, Italian, Spanish, Basque, Chinese and Japanese and the domain is the environment. However, the system can be extended for any other language and to any other domain.

Whereas the current Wikipedia uses free text to share knowledge, KYOTO will represent this knowledge so that a computer can understand it. For example, the notion of environmental footprint may become defined in the same way in all these languages but also in such a way that the computer knows what information is necessary to calculate a footprint. With these definitions it will be possible to find information on footprints in documents, websites and reports so that users can directly ask the computer for actual information in their environment, e.g. what is the footprint of their town, their region, their company or their personal footprint.

The terms for a community are automatically acquired from the documents in the domains but can be edited and modified by the domain experts in the Wiki. The edited terms are represented as a domain wordnet that is linked to generic wordnets in the languages. Similarly, a domain ontology is created that is linked to a generic ontology. The editing is supported through text from which the terms are derived and complex knowledge structures are hidden for the user. The KYOTO platform is thus not developed for knowledge engineers but for Semantic Web2.0 communities that maintain their own knowledge and wordnets, creating Semantic Web3.0 structures and services. By linking domain wordnets to generic wordnets and to a shared ontology, the KYOTO project can become the backbone for developing the Global Wordnet Grid.

When the domain modeling is done, the acquired knowledge can be used for open text mining. Users can specify a conceptual profile that is mapped to the text in the domain through expression rules and wordnets to find fact in the text. The facts are aggregated from the text and presented in an RDF repository for search. The conceptual knowledge in the community is thus directly applied for interpretation of the factual knowledge in the community.

Horacio Rodríguez, Universitat Politècnica de Catalunya. Miércoles, 9 de septiembre. 11:00-12:00

Conocimiento rico, conocimiento pobre y desconocimiento en PLN. El ejemplo de la implicación textual.

En los últimos tiempos proliferan en el área del tratamiento de la lengua, y en muchas otras disciplinas que manejan conocimiento humano, los sistemas que se autocalifican de basados en conocimiento pobre, ligero, superficial, etc. A veces la tarea de procesamiento lingüístico que se aborda no requiere un conocimiento más profundo para lograr sus objetivos, a veces la adquisición y uso de conocimiento más rico es simplemente difícil o cara o lenta, a veces no somos capaces de extraer o utilizar el conocimiento que sería más apropiado, a veces simplemente ignoramos cuál podría ser el conocimiento más apropiado para abordar la tarea.

El uso de un tipo u otro de conocimiento va obviamente ligado a la aproximación que se utilice para resolver la tarea planteada. Los sistemas basados en aprendizaje automático o los estadísticos son más robustos frente a las limitaciones del conocimiento a emplear. Cuando se habla de conocimiento no debemos restringirnos al conocimiento lingüístico, el conocimiento básico, el conocimiento de sentido común o el conocimiento del mundo juegan un papel relevante en nuestros procesos.

La conferencia abordará estos temas en el marco del PLN y concretamente en una aplicación, la implicación textual, especialmente sensible al tipo y nivel del conocimiento a emplear (o a su ausencia).

Joakim Nivre, Uppsala University. Thursday, 10th of September. 11:00-12:00

Current Trends in Data-Driven Dependency Parsing.

Dependency-based syntactic parsing has become a standard technique in natural language processing and a number of different models have been proposed in recent years, in particular data-driven models that can be trained using syntactically annotated corpora, or treebanks. Most of these models can be characterized as either graph-based or transition-based. Graph-based models learn to score entire dependency trees and use exact search to find the best tree for a given input sentence. Transition-based models learn to score local parsing actions and use greedy search to find the best sequence of actions for a given input sentence. Both types of models give state-of-the-art accuracy but a comparative error analysis reveals that they have different error distributions and that this difference can be tied to theoretical properties of the models. Recent work on data-driven parsing has therefore to a large extent been characterized by attempts to combine the strengths of the two models, either through the development of hybrid models or through system combination. In this talk, I review the classic graph-based and transition-based models, characterize their typical strengths and weaknesses, and report on recent work aiming to improve the basic models.