

Generación de Lenguaje Natural

Año 2011

Pablo Ariel Duboue

2 de diciembre de 2010

1. Generación de Lenguaje Natural (NLG)

Texto principal: Building Natural Language Generation Systems (2000)
Reiter & Dale (ISBN 0521620368)

- a) Introducción a la generación de lenguaje natural. Arquitectura de sistemas de generación. Ejemplos de sistemas de generación. Datos de entrada a un sistema de generación de lenguaje natural.

Papers:

- 1) E. Reiter, Has a Consensus NLG Architecture Appeared, and is it Psycholinguistically Plausible?, IWNLG '94.
- b) Planeamiento de contenidos. Selección de contenidos. Selección de contenidos por spreading activation. Ordenamiento de contenidos. Conocimiento comunicacional específico (Domain Communicative Knowledge). Métodos estadísticos de selección y ordenamiento de contenidos.

Papers:

- 1) Cox & al., '99. Dynamic versus static hypermedia in museum education: an evaluation of ILEX, the intelligent labelling explorer.
- 2) Kittredge & al, '91. On the need for domain communication language.
- 3) Duboue & McKeown '03, Statistical Acquisition of Content Selection Rules for Natural Language Generation.
- 4) Barzilay & Lapata, '08. Modeling Local Coherence: An Entity-based Approach.
- c) Planeamiento de oraciones. Lexicalización de conceptos. Agrupamiento de oraciones. Pronominalización. Generación de expresiones referentes. Métodos estadísticos de generación de expresiones referentes.

Papers:

- 1) Reape & Mellish, '98. Just What is Aggregation Anyway?
 - 2) Dale, '09. Referring expression generation through attribute-based heuristics.
 - 3) Belz, '08. Intrinsic vs. extrinsic evaluation measures for referring expression generation.
- d) Generación de oraciones. Métodos por unificación de gramáticas funcionales. FUF. Métodos estadísticos de generación de oraciones. Nitrogen. SPUD.

Papers:

- 1) Elhadad & Robin, '99. SURGE: a comprehensive plug-in syntactic realization component for text generation.
 - 2) Langkilde & Knight, '98. The practical value of n-grams in generation.
 - 3) Stone & al, '03. Microplanning with communicative intentions: The SPUD system.
- e) Generación multi-lenguaje. Re-uso de componentes. Evaluación de sistemas de generación. Clinical trials. Gold standards.

Papers:

- 1) Callaway, '05. Automatic cinematography and multilingual NLG for generating video documentaries.
- 2) Reiter & Robertson, '03. Lessons from a failure: Generating tailored smoking cessation letters.
- 3) Reiter, '02. Should corpora texts be gold standards for NLG?

2. Resumen Automático.

Texto principal: Advances in Automatic Text Summarization (1999). Mani & Maybury ISBN 0262133598.

- a) Introducción al resumen automático de textos. Resumen automático mono-documento vs. multi-documento. Ejemplos de sistemas de resumen automático.
- b) Resumen automático mono-documento. Sistemas basados en conocimiento. Sistemas estadísticos. Ultra-sumarización. Generación de títulos.

Papers:

- 1) Silber & McCoy, '02. Efficiently Computed Lexical Chains as an Intermediate Representation for Automatic Text Summarization.
- 2) Mittal, '99. Ultra-Summarization: A Statistical Approach to Generating Highly Condensed Non-Extractive Summaries.
- c) Resumen automático multi-documento. Sistemas extractivos. Sistemas basados en fusión de oraciones. Resumen de documentos en múltiples lenguajes. Actualización de resúmenes.

Papers:

- 1) Radev & al. '04. Centroid-based summarization of multiple documents.
- 2) Barzilay, '05. Sentence fusion for multidocument news summarization.
- 3) Saggion & al, '02. Developing infrastructure for the evaluation of single and multi-document summarization systems in a cross-lingual environment.
- d) Evaluación de resúmenes. Evaluación intrínseca y extrínseca. Método de evaluación basado en pirámides de contenidos. ROUGE.

Papers:

- 1) Dorr & al, '05. A methodology for extrinsic evaluation of text summarization: does ROUGE correlate?
Lin, '04. Rouge: A package for automatic evaluation of summaries.
Nenkova & al, '07. The Pyramid Method: Incorporating human content selection variation in summarization evaluation.

3. Traducción Automática.

Texto principal: Statistical Machine Translation (2010). Koehn (ISBN: 0521874157)

- a) Introducción a la traducción automática. Producción de textos durante la traducción automática.
- b) Métodos basados en el conocimiento. Reglas de transferencia. Gramáticas bi-direccionales.

Papers:

- 1) Lavoie & al, '00. A framework for MT and multilingual NLG systems based on uniform lexico-structural processing.
 - 2) Wu, '97. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora.
- c) Métodos estadísticos. Decodificación estadística. Búsqueda por haces.

Papers:

- 1) Al-Onaizan & al, '99. Giza: Statistical machine translation.
 - 2) Koehn, '04. Pharaoh: a beam search decoder for phrase-based statistical machine translation models.
- d) Evaluación de traducción automática. BLEU.

Papers: Papinini & al '01. Bleu: a Method for Automatic Evaluation of Machine Translation.