

Universidad de La Habana
Facultad de Matemática y Computación



Sistema para la generación de resúmenes textuales de eventos deportivos independientes de la fuente de datos.

Autor:

Abel Molina Sánchez

Tutores:

Dr. Yudivián Almeida Cruz

Lic. Manuel Santiago Fernández Arias

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencia de la Computación

Fecha

github.com/abel1927/Thesis

Dedicación

Agradecimientos

Agradecimientos

Opinión del tutor

Opiniones de los tutores

Resumen

Resumen en español

Abstract

Resumen en inglés

Índice general

Introducción	1
1. Estado del Arte	5
1.1. Determinación del contenido	5
1.2. Estructuración del texto	6
1.3. Agregación	7
1.4. Lexicalización	8
1.5. Expresiones de referencia	9
1.6. Realización lingüística	10
2. Propuesta	11
3. Detalles de Implementación y Experimentos	12
Conclusiones	13
Recomendaciones	14
Bibliografía	15

Índice de figuras

Ejemplos de código

Introducción

Desde hace varios años la Inteligencia Artificial(IA) viene revolucionando e impactando significativamente en muchas esferas de la vida de la vida del hombre. Dentro de la IA hay muchos campos de estudios y entre ellos el Procesamiento de Lenguaje Natural(PLN) es el encargado de lidiar con las problemáticas relacionadas con el lenguaje humano. Uno de los subcampos dentro del PLN que está acaparando especial atención en los últimos tiempos es el de la Generación de Lenguaje Natural(GLN) y en especial aquel que está enfocado a expresar a través de texto información contenida en forma de datos no lingüísticos(D2T por sus siglas en inglés, Data-to-Text). Existe consenso en la forma en que la salida de un sistema de GLN debe presentarse: texto, valga la redundancia, pero no hay un estándar en cuanto a la forma en que se presentan los datos para su procesamiento, variando de un sistema a otro. Como regla general el texto producido por estos sistemas debe mantener fidelidad a los datos que lo originan, así como debe ser consecuente con su intención comunicativa, no siendo lo mismo un sistema para la generación de diálogos que uno que tiene como objetivo describir resúmenes biográficos.

En un contexto donde la producción de datos se ha acelerado resultado de los avances tecnológicos y de los sistemas industriales, se ha hecho necesario para las empresas el manejo y la interpretación de estos datos. Empresas como *Automated Insights*¹ se han especializado en la GLN y han enfocado su negocio en brindar soluciones a otras corporaciones en vista a automatizar sus procesos de producción de texto. Entre los casos de uso clásico de estas soluciones encontramos, en el marco del comercio electrónico, la generación de descripciones de productos a partir de sus fichas técnicas. El periodismo ha sido otra de las esferas beneficiadas de estos avances en la GLN. La generación automática de noticias(conocida como periodismo robótico) está cada vez más extendida, al punto de que importantes editoriales como *The Washington Post* han creado sus propios sistemas para la generación de texto a partir de datos². En este caso su sistema se apoda *Heliograf* y les permite cubrir todos los partidos de fútbol americano de las escuelas secundarias del área de Washington

¹<https://automatedinsights.com/>

²<https://www.washingtonpost.com/pr/wp/2017/09/01/the-washington-post-leverages-heliograf-to-cover-high-school-football/>

DC cada semana.

Por su carácter estadístico y su gran audiencia, el deporte es una de las esferas que brinda grandes posibilidades para desarrollar sistemas de generación automática de texto. Es relativamente sencillo encontrar en la web los datos o tablas estadísticas de un determinado enfrentamiento deportivo, pero la gran variedad de eventos que se suceden constantemente hacen que sea humanamente imposible darle coberturas a cada uno de ellos a nivel de narración y resúmenes. A partir de lo cual surgen soluciones que cubren determinados eventos de interés como el ejemplo mencionado en el párrafo anterior.

Problema

Aún con el desarrollo de los sistemas de GLN y su mayor asimilación en diversos ámbitos, siguen siendo absolutamente predominantes los sistemas que tienen el inglés como principal o único lenguaje de referencia. El lenguaje español no presenta tanta cobertura en este campo. De la misma forma esto impacta directamente en los sistemas que tienen como objetivo comunicativo los eventos deportivos. Siendo el deporte un objeto de mucho interés en la comunidad hispanohablante en general y en Cuba en particular, se hace necesario ampliar los escenarios presentes desde esta perspectiva.

Los sistemas presentes en la literatura en su mayoría se basan en un conocimiento explícito del dominio a tratar así como en una estructuración predefinida de los datos en base al dominio. A su vez, son muchas las propuestas que parten de la obtención de los datos desde la fuente como parte propia del sistema y no propiciados por el usuario. Son pocos los sistemas funcionales, fuera de la industria, capaces de desacoplarse de las fuentes de datos y de abarcar desde una misma estructura distintos dominios.

Los recientes modelos neuronales están impactando el estado del arte de muchas tareas en el campo del PLN. Se hace necesario acercar el prisma hacia los mismos desde el punto de vista de la GLN.

Motivaciones

La propuesta del presente trabajo está motivada en las problemáticas anteriores. Siendo el deporte un campo que despierta tanto interés y que es fuente de entretenimiento de tantas personas, sentar las bases de una metodología para la generación de resúmenes de eventos deportivos es un reto estimulante.

En el presente trabajo se presenta un sistema capaz de generar resúmenes de eventos deportivos a partir de los datos proveídos por el usuario. El mismo servirá, dado el caso, para dar descripción a eventos de interés de un determinado público que pudieran quedar fuera de la cobertura noticiosa de los medios de prensa.

Así mismo este trabajo se presenta en el marco de las líneas de investigación exis-

tentes en el grupo de Inteligencia Artificial de la Facultad de Matemática y Computación de la Universidad de La Habana(MATCOM). Servirá como base para futuros trabajos que sigan ampliando y profundizando en la GLN.

Antecedentes

Esta vía investigativa dentro del departamento de IA de MATCOM comenzó con la propuesta en 2019 de realizar un modelo GLN capaz de dar cobertura a las actuaciones destacadas de los peloteros cubanos en la Grandes Ligas de Béisbol. Dicha propuesta derivó en el trabajo de tesis de Roberto Balboa González(Balboa 2020) que sirve de antecedente del presente trabajo.

Objetivo

El objetivo general de este trabajo es proponer una metodología funcional para la generación de resúmenes de enfrentamientos deportivos independientes de la fuente de datos del dominio. A partir de dicha metodología se presenta la implementación de un sistema capaz de crear dichos resúmenes en idioma español. Para la consecución del objetivo es necesario:

- Estudiar detalladamente la literatura relacionada con la GLN y determinar los enfoques factibles.
- Comprobar los sistemas de generación relacionados con el dominio(enfrentamientos deportivos) y sus características.
- Estudiar en profundidad el dominio, y determinar los rasgos distintivos y comunes entre los eventos deportivos.
- Diseñar un esquema de eventos comunes característicos entre los deportes que de lugar a un esquema específico de cada uno.
- Plantear una estructura intermedia para el ingreso de los datos en base al esquema general diseñado.
- Diseñar e implementar los modelos generadores de texto.
- Presentar el sistema funcional para la generación de resúmenes.

Estructura del trabajo

El resto del trabajo se organiza de la siguiente manera. El capítulo 1 aborda los problemas generales de los sistemas de GLN y las distintas técnicas relevantes en

la literatura para su solución. En el capítulo 2 se presenta el diseño del modelo a implementar, mientras que en el capítulo 3 se analizan los detalles de la implementación y los resultados. El trabajo concluye con la presentación de las conclusiones y recomendaciones para futuros trabajos.

Capítulo 1

Estado del Arte

Hasta hoy considerado como el punto de referencia y texto más completo en el campo de GLN(Gatt y Krahmer 2018), el libro de Ehud Riter y Robert Dale: Building Natural Language Generation Systems(Reiter y Robert Dale 2000), sienta las bases y define las que se han asumido como tareas principales de un sistema de GLN:

- Determinación del contenido: decidir qué información es relevante.
- Estructuración del texto: determinar en qué orden se presentará la información.
- Agregación: decidir qué información se presenta en cada oración.
- Lexicalización: encontrar las palabras y frases adecuadas para expresar información.
- Generación de expresiones de referencia: selección de las palabras y frases para identificar entidades del dominio.
- Realización lingüística: conformación del texto con oraciones bien formadas.

Cada una de estas tareas presenta un objetivo dentro del sistema y existen distintas técnicas para su realización.

1.1. Determinación del contenido

Cada sistema GLN tiene un objetivo comunicativo, lo cual hace necesario determinar qué información del dominio es relevante y debe estar representada en el texto final(Reiter y Robert Dale 2000). Los sistemas pueden presentar una especificación de los datos de la entrada o seleccionar un subconjunto de los mismos(Reiter y Robert Dale 2000).

El receonocimiento de patrones es una de las técnicas utilizadas con este fin. SumTime-Turbine(Yu y col. 2006) es un sistema funcional para crear informes sobre un mecanismo de turbinas de gas. En un contexto donde la cantidad de datos es muy grande y en su mayoría irrelevante, los autores con ayuda de los expertos del dominio crean un modelo que permite reconocer patrones en los datos así como luego, utilizando una base de datos de patrones antiguos determinar los patrones relevantes a ser expresados en el informe.

Existen muchos sistemas de GLN que utilizan modelos basados en reglas como única metodología para la selección de contenido(Perera y Nand 2017; Reiter y Robert Dale 2000). Bouayad-Agha, Casamayor y Wanner (Bouayad-Agha y col. 2011) en su trabajo presentaron el diseño teórico de un sistema para generar resúmenes de partidos de fútbol de la Liga Española. Para la conformación del mismo describieron la construcción de una gran base de conocimientos del dominio así como un estricto sistema de reglas para la selección de contenido. Es relevante este enfoque ya que con un conocimiento del dominio a tratar y de la intención comunicativa del texto a producir es posible implementar reglas que den lugar a la selección del contenido relevante(Reiter y R. Dale 1997; Reiter y Robert Dale 2000).

Como el proceso de creación de reglas puede ser tedioso y es dependiente del dominio(Reiter y R. Dale 1997) aparecieron iniciativas que para la automatización de esta tarea. Un enfoque basado en técnicas de aprendizaje automático es el presentado por Duboué y McKeown(Duboué y K. McKeown 2003). En este trabajo plantearon un modelo de aprendizaje supervisado que a través de un corpus de resultados deseados(texto escrito por humanos) aparejados con los tipos lingüísticos de la entrada(distintos tipos de datos que recibe el sistema) determina un conjunto de constantes. Estas constantes expresan si determinado dato de entrada debe aparecer o no reflejado en la salida y bajo que condiciones. Este sistema se utilizó para la generación de descripciones biográficas cortas que resumirán hechos importantes sobre personajes famosos.

1.2. Estructuración del texto

La estructuración del texto o planificación del discurso es el proceso donde se da orden y estructura al conjunto de mensajes a expresar en el texto producido. En un texto la información se presenta en un orden particular y, por lo general, hay una estructura subyacente a la presentación. La complejidad de la estructura de los textos puede variar de un sistema a otro. Una buena estructuración puede hacer que un texto sea mucho más fácil de leer(Reiter y R. Dale 1997).

Los primeros enfoques para la estructuración de documento se basaron en reglas estructuradas hechas a mano dependientes del dominio(Gatt y Krahmer 2018). A este enfoque se le conoce como el enfoque basado en esquemas, nombre que se derivó

del trabajo de Kathleen R. McKeown(K. R. McKeown 1985) cuando acuñó el término "*schematta*". En la construcción de su sistema TEXT(K. R. McKeown 1985), McKeown, luego del análisis de muchos ejemplos del dominio, concluyó que dado un objetivo comunicativo, la información tiende a transmitirse en el mismo orden. En base a esto definió estructuras(esquemas) que determinan posibles combinaciones de atributos, formando patrones y plantillas. De esta forma el sistema, dada una intención comunicativa, puede seleccionar un esquema que defina la forma de transmitir la información. La mayoría de los sistemas que siguen este enfoque utilizan las Matrices de Valores de Atributos(AVMs por sus siglas en inglés, Attribute Value Matrics)(Perera y Nand 2017).

Las estructuras retóricas son otro de los mecanismos que se utilizan para la planificación. Estas se derivan del trabajo de Mann y Thompson quienes introdujeron la Teoría de la Estructura Retórica(RST por sus siglas en inglés, Rhetorical Structure Theory)(Mann y Thompson 1988). Las estructuras retóricas constituyen un método lingüístico para la descripción de texto caracterizando las estructuras primarias del mismo y estableciendo relaciones funcionales entre sus distintas partes. La RST tiene como base los conceptos de núcleo y satélite que definen las partes del texto entre las que se establece una relación. Los distintos sistemas que utilizan las estructuras retóricas para la estructuración del texto definen que tipo de relaciones establecen. Ejemplo de relaciones lingüísticas planteadas por Mann y Thompson en su trabajo son: motivación, causa, condición, circunstancia, etc(Mann y Thompson 1988).

1.3. Agregación

En un texto, cada parte de información no tiene que estar presente en oraciones independientes. Hay escenarios donde es deseable que distintos mensajes sean transmitidos en una misma oración. La agregación puede permitir crear texto de mayor calidad o eliminar repeticiones innecesarias que vayan en contra de la fluidez del texto(Gatt y Krahmer 2018). Un ejemplo de agregación, en el dominio del fútbol, describiendo el hecho de dos anotaciones consecutivas de un jugador, pudiera ser:

- (1) Ronaldo anotó para el Real Madrid en el minuto 2. Ronaldo anotó para el Real Madrid en el minuto 8.
- (2) Ronaldo anotó dos veces para el Real Madrid antes del minuto 8.

En (2) se evita la repetición y la información se presenta de forma más fluida y natural al lector.

Reape y Mellish(Reape y Mellish 1999) realizaron un estudio sobre la tarea de agregación dentro de los sistemas de generación de texto, distinguiendo entre la agregación a nivel semántico(más dependiente del dominio) y a nivel sintáctico. Muchos

de los primeros trabajos sobre agregación fueron dependientes del dominio, centrados en la aplicación de reglas (por ejemplo, "si un jugador marca dos goles consecutivos, exprésalo en la misma frase") generalmente hechas a mano (ejemplo Shaw 1998). Con el tiempo aparecieron propuestas que utilizan enfoques de aprendizaje automático. SPoT es un sistema que se presentó en Walker y col. 2001) constituyendo uno de los primeros sistemas entrenados que incluye la tarea de agregación. Los autores plantearon una metodología basada en la producción de varios textos para una misma entidad informativa utilizando diferentes cláusulas de agregación asociadas al dominio. Después utilizaron un modelo entrenado para dar un valor a cada una de las salidas estableciendo un ranking a partir del cual hacer la selección. Mientras, Barzilay y Lapata (Barzilay y Lapata 2006) plantearon el problema en términos de optimización global. Realizan una clasificación inicial sobre pares de entradas de la base de datos que determina si deben agregarse o no en función de su similitud por pares. Posteriormente, seleccionan un conjunto globalmente óptimo de entradas relacionadas en función de un grupo de restricciones.

Con la agregación sintáctica, podría decirse que es más factible definir reglas independientes del dominio para eliminar la redundancia (Gatt y Krahmer 2018). Por ejemplo, convertir (3) en (4) a continuación

- (3) Ronaldo marcó en el minuto 2 y marcó de nuevo en el minuto 8.
- (4) Ronaldo marcó en el minuto 2 y de nuevo en el 8.

podría lograrse identificando las frases verbales paralelas en las dos oraciones conjuntas y eliminando el sujeto y el verbo en la segunda.

1.4. Lexicalización

La lexicalización es un proceso muy importante dentro de un sistema de GLN. Es el proceso durante el cual se seleccionan la palabra o palabras que expresan un concepto o relación (Reiter y R. Dale 1997). Una de las complicaciones del proceso de lexicalización está dada porque una misma relación puede ser expresada de distintas formas. Por ejemplo, el evento de la anotación de un gol en un partido de fútbol puede ser expresado como: "marcar un gol", "poner el balón en la red", "conseguir una anotación". La complejidad de este proceso depende en gran medida del número de alternativas que el sistema pueda o quiera contemplar. Las restricciones contextuales también juegan un papel importante a la hora de expresar un mensaje, por ejemplo, la expresión "marcó un gol" es desafortunada si el evento descrito es un gol en propia puerta (Gatt y Krahmer 2018).

El proceso de lexicalización puede seguir dos vertientes principales. Una sería la realización de la lexicalización de la forma más simple posible, lo cual se lleva a

cabo generalmente utilizando técnicas para el llenado de plantillas lexicalizadas. De otra forma se puede realizar este proceso en mayor profundidad utilizando técnicas más complejas que permitan por ejemplo: la eliminación de palabras innecesarias, la selección de vocablos que maximicen la efectividad del objetivo comunicativo del texto, la unión de términos que se aparejan frecuentemente en el dominio(Perera y Nand 2017).

Los enfoques basados en reglas son de los más utilizados, pudiendo variar en complejidad entre un sistema y otro. *EasyText*(Danlos y col. 2011) es un ejemplo de sistema que utiliza reglas para la lexicalización pero que da un paso más allá ya que consume de una base de datos léxica creada principalmente por lingüistas. Esta alternativa permite una lexicalización más avanzada pero a su vez es altamente costosa en recursos.

La utilización de ontologías también está presente en este proceso de los sistemas de GLN. El uso de ontologías permite al sistema ganar en adaptabilidad ya que encontrar una ontología para un dominio determinado es más sencillo que encontrar un corpus para el mismo. Así mismo ofrecen una mayor cobertura de las representaciones semánticas que los corpus(Perera y Nand 2017). En Cimiano y col. 2013 introducen un modelo que utiliza este enfoque presentado en el dominio de las recetas de cocina.

1.5. Expresiones de referencia

Robert Dale y Ehud Rither describieron la generación de expresiones de referencias dentro de un sistema de GLN como la tarea de indentificar la expresión, comprensible de cara al usuario, a utilizar para indentificar a una instancia del dominio(Gatt y Krahmer 2018; Reiter y Robert Dale 2000). Los primeros métodos para la selección de referencias fueron los algoritmos generativos que en común presentan la necesidad de tener conocimiento contextual y de propiedades de las entidades(Gatt y Krahmer 2018). De este orden es el algoritmo incremental cuya base se plantó en Robert Dale y Reiter 1995. El algoritmo, en base a conocer la entidad a referenciar(objetivo), el resto de entidades(llamadas distractores) y el grupo de propiedades que definen entidades en el dominio, busca determinar un conjunto de propiedades únicas que definan al objetivo y lo diferencien del resto.

Muchos de los trabajos que versan sobre este tema hacen énfasis en un tipo determinado de referencia(Ferreira y col. 2018). La selección puede ser de un pronombre(él/ella), una descripción(Simón, el jugador cubano) o en la generación de nombres propios(Frederich Cepeda/Cepeda). Rither y Dale(Reiter y Robert Dale 2000) plantean una diferencia entre una referencia temprana(primer vez que se menciona una entidad en el texto) y una tardía(cuando se refiera a una entidad mencionada anteriormente). Plantearon el uso de los nombres propios a la hora de introducir una

entidad, luego, en base a cláusulas seleccionar un pronombre apropiado, ejemplo:

```
si el referente fue mencionado en la oración anterior;
entonces utiliza un pronombre
```

a su vez es necesario considerar los escenarios donde la generación de expresiones de referencia pudiera llevar a ambigüedades:

- Benzema anotó dos para el Real Madrid mientras Luka Modric brindó dos asistencias. Él fue elegido el jugador del partido...

En Siddharthan y col. 2011 los autores realizaron un estudio empírico del comportamiento de las referencias hacia personas basadas en su nombre propio en el contexto de los artículos de noticias. Para ello utilizaron un corpus de noticias en inglés de diferentes agencias de prensa y cuantificaron las diferentes formas de referencia según el momento referencial de la instancia (temprana o tardía). Como resultado de este trabajo arrojaron que el nombre completo de la entidad suele utilizarse como primera referencia en la práctica totalidad de los casos mientras que en su mayoría el apellido se utiliza cuando se trata de una referencia tardía.

1.6. Realización lingüística

El proceso de realización lingüística es el que da lugar a la formación final del texto expresado en oraciones con una estructura gramatical correcta y coherente con el mensaje a transmitir. Esta tarea implica ordenar los constituyentes de una oración, así como generar las formas morfológicas correctas (incluidas las conjugaciones y la concordancia de los verbos). A menudo, los realizadores también necesitan insertar palabras funcionales (como verbos auxiliares y preposiciones) y signos de puntuación (Gatt y Krahmer 2018).

Una de las técnicas más utilizadas es la que incluye el uso de plantillas predefinidas para expresar mensajes (Gatt y Krahmer 2018). Las plantillas, aunque requieren una carga intensiva de trabajo para lograr mayor variabilidad en el texto a producir permiten un control total sobre la calidad y la correctitud del texto a elaborar. Un ejemplo de plantilla:

```
$equipo_1 venció $pts_equipo_1 a $pts_equipo_2 a $equipo_2
```

Dicha plantilla, que representa el resultado de un enfrentamiento entre dos equipos se completa con la información extraída de los datos y a la hora de la realización el resultado pudiera ser el siguiente:

```
Industriales venció 10 a 1 a Granma
```

Capítulo 2

Propuesta

Capítulo 3

Detalles de Implementación y Experimentos

Conclusiones

Conclusiones

Recomendaciones

Recomendaciones

Bibliografía

- Balboa, R. (2020). Generación de noticias a partir de datos estructurados en el dominio del béisbol (vid. pág. 3).
- Barzilay, R. & Lapata, M. (2006). Aggregation via Set Partitioning for Natural Language Generation. *NAACL* (vid. pág. 8).
- Bouayad-Agha, N., Casamayor, G. & Wanner, L. (2011). Content selection from an ontology-based knowledge base for the generation of football summaries. *ENLG* (vid. pág. 6).
- Cimiano, P., Lüker, J., Nagel, D. & Unger, C. (2013). Exploiting ontology lexica for generating natural language texts from RDF data. *Proceedings of the 14th European Workshop on Natural Language Generation*, 10-19 (vid. pág. 9).
- Dale, R. [Robert] & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2), 233-263 (vid. pág. 9).
- Danlos, L., Meunier, F. & Combet, V. (2011). EasyText: an operational NLG system. *ENLG 2011-13th European Workshop on Natural Language Generation* (vid. pág. 9).
- Duboué, P. A. & McKeown, K. (2003). Statistical Acquisition of Content Selection Rules for Natural Language Generation. *EMNLP* (vid. pág. 6).
- Ferreira, T. C., Moussallem, D., Kádár, A., Wubben, S. & Krahmer, E. (2018). NeuralREG: An end-to-end approach to referring expression generation. *arXiv preprint arXiv:1805.08093* (vid. pág. 9).
- Gatt, A. & Krahmer, E. J. (2018). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. *ArXiv, abs/1703.09902* (vid. págs. 5-10).
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3), 243-281 (vid. pág. 7).
- McKeown, K. R. (1985). Discourse strategies for generating natural-language text. *Artificial intelligence*, 27(1), 1-41 (vid. pág. 7).

- Perera, R. & Nand, P. (2017). Recent Advances in Natural Language Generation: A Survey and Classification of the Empirical Literature. *Comput. Informatics*, 36, 1-32 (vid. págs. 6, 7, 9).
- Reape, M. & Mellish, C. (1999). Just what is aggregation anyway. *Proceedings of the 7th European Workshop on Natural Language Generation*, 20-29 (vid. pág. 7).
- Reiter, E. & Dale, R. [R.]. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3, 57-87 (vid. págs. 6, 8).
- Reiter, E. & Dale, R. [Robert]. (2000). *Building Natural Language Generation Systems*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519857>. (Vid. págs. 5, 6, 9)
- Shaw, J. (1998). Clause Aggregation Using Linguistic Knowledge. *INLG* (vid. pág. 8).
- Siddharthan, A., Nenkova, A. & McKeown, K. (2011). Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4), 811-842 (vid. pág. 10).
- Walker, M., Rambow, O. & Rogati, M. (2001). SPoT: A trainable sentence planner. *Second Meeting of the North American Chapter of the Association for Computational Linguistics* (vid. pág. 8).
- Yu, J., Reiter, E., Hunter, J. & Mellish, C. (2006). Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering*, 13, 25-49 (vid. pág. 6).