



Taller 3 – 28 de septiembre, 17:00 a 20:00

Uso de extractores automáticos de terminología para el análisis textual

Tallerista: Sahara Iveth Carreño Cruz, Departamento de Lenguas, Universidad Católica de Temuco, scarreno@uct.cl

Objetivo: Conocer el funcionamiento básico de los extractores terminológicos para determinar su utilidad en la explotación de corpus textuales.

Público al que está dirigido: investigadores en proceso de analizar corpus textuales que tengan poco conocimiento respecto a herramientas de explotación de tales recursos.

Cupos: 15 participantes

Resumen

En diversos estudios lingüísticos basados en corpus textuales, las unidades que cobran particular relevancia son las lexías simples y los frasemas, esto es, las combinaciones léxicas de diferente naturaleza y grado de cohesión. La información extraída a partir de estas unidades facilita y apoya en el análisis de los textos en diversos planos, por ejemplo, semántico, conceptual, temático, discursivo, de grado de especialidad, por mencionar algunos.

En efecto, en campos como la lingüística aplicada, la terminología y la traducción existen estudios que recurren a la extracción terminológica como parte de su metodología. En los trabajos de Bowker (2000); Bowker y Pearson (2002); Cabré et al. (2002); Warburton (2015), entre otros, se describe este tipo de herramientas y su aplicabilidad en distintos ámbitos.

Por tanto, proponemos una sesión en la que se muestren al menos dos extractores terminológicos de los que se dispone en la Web y que se apliquen a un conjunto de textos de manera que los participantes determinen el potencial de uso de los mismos para sus propias investigaciones.

El taller consta de cuatro etapas principales:

- a) Contextualización – Se explicarán algunos conceptos clave y el tipo de sistemas de extracción que existen en la actualidad. Se presentarán brevemente algunos ejemplos de estudios que recurren a la extracción automática de terminología y, por último, nos centraremos en la descripción del funcionamiento general de los sistemas de extracción híbridos (lingüístico-estadísticos) que existen para lenguas modernas. Esta etapa tendrá una duración aproximada de 80 minutos y comprenderá una sesión de consultas.



- b) Tutorial con dos extractores monolingües – Se han preseleccionado sistemas como SketchEngine, TermoStat, TerMine y YATE. Su selección final en el momento del taller dependerá de las lenguas de trabajo de los participantes. Se aplicará un sistema a un corpus de textos en una lengua extranjera y un segundo en un corpus en castellano. Se analizarán los resultados arrojados por cada extractor, esto es, lista de candidatos a términos, y otras informaciones lingüísticas y estadísticas presentadas, tales como combinaciones recurrentes, índice de pertinencia, frecuencia, distribución en nube, estructuras sintagmáticas, candidatos en contexto, entre otros. También se mostrará cómo exportar los datos (en hoja de cálculo, por ejemplo), más una herramienta para consultar los candidatos en contexto (por ejemplo, con un programa de concordancias independiente como TextStat). Esta etapa tendrá una duración de máxima de 80 minutos.
- c) Aplicación en colecciones individuales de textos – Los participantes podrán aplicar el sistema de extracción que deseen sobre sus propias colecciones de textos. Esta actividad permitirá el trabajo más individualizado y experimental. Se dedicarán aproximadamente 25 minutos a esta tarea.
- d) Discusión de usos potenciales y conclusión – En conjunto, y con base en los resultados evidenciados, se analizarán las ventajas y limitaciones de los extractores terminológicos y se discutirán los usos potenciales de estos sistemas en contextos específicos de investigación. Esta fase final tendrá una duración máxima de 25 minutos.

Bibliografía

- Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: Presses de l'Université d'Ottawa.
- Bowker, Lynne et Pearson, Jennifer. 2002. *Working with Specialized Language- a practical guide to using corpora*. London/ New York: Routledge.
- Cabré, M. T., Estopà, R. y Vivaldi, J. (2001). "Automatic term detection: A review of current systems". En D. Bourigault, C. Jacquemin, y M.C. L'Homme (ed.), *Recent advances in Computational Terminology*, Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 53-87.
- Warburton, K. (2015). "Terminology Management". En Ch. Sin-wai. *The Routledge Encyclopedia of Translation Technology*. London/Nueva York: Routledge.

Reseña biográfica

Sahara Iveth Carreño Cruz es traductora y terminóloga de español, inglés y francés. Sus áreas de investigación son la terminología, la adquisición automática de terminología y la traducción automática y traducción asistida. Magíster en Traducción por la Universidad de Montreal, Canadá.



Desde septiembre de 2010 trabaja como docente del Departamento de Lenguas y Traducción de la Universidad Católica de Temuco e imparte las asignaturas relativas a las tecnologías de la traducción y la terminología.