

Sistemas de Recuperación de Información Geográfica multilingües en CLEF*

Multilingual Geographical Information Retrieval systems in CLEF

José Manuel Perea Ortega
Manuel García Vega

Miguel Angel García Cumbreiras
L. Alfonso Ureña López

Universidad de Jaén, Campus Las Lagunillas
Edificio A3. E-23071

{jmperea,magc,mgarcia,laurena}@ujaen.es

Resumen: En este artículo se presenta un estudio comparativo de las distintas estrategias y técnicas de procesamiento del lenguaje natural más utilizadas en la actualidad para abordar la tarea de la recuperación de información geográfica (*Geographical Information Retrieval*, GIR). Este trabajo se ha basado fundamentalmente en el análisis de los mejores sistemas presentados a la tarea de búsqueda del GeoCLEF, un marco de evaluación para recuperación de información geográfica que pertenece al foro internacional *Cross Language Evaluation Forum* (CLEF). Las conclusiones obtenidas reflejan que es imprescindible hacer uso de recursos externos de información geográfica, tales como *gazetteers* y tesauros o reconocedores de entidades. Así mismo es necesario realizar una indexación por separado de la información geográfica y de la no geográfica antes del proceso de recuperación.

Palabras clave: Recuperación de Información Geográfica, GeoCLEF, Procesamiento del Lenguaje Natural, Recuperación de Información

Abstract: This paper presents a comparative study of several strategies and techniques of natural language processing most used at present to solve the geographical retrieval information (GIR) task. This work has been based on the analysis of the best systems submitted to the search task of GeoCLEF, an evaluation framework for the geographical information retrieval task which belongs to the international forum Cross Language Evaluation Forum (CLEF). The main conclusions show that it is imperative to make use of external geographic information resources such as gazetteers and thesaurus, named entity recognizers and it is necessary to make an index for geographic information only and another index for non-geographic information before the retrieval process.

Keywords: Geographical Information Retrieval, GeoCLEF, Natural Language Processing, Information Retrieval

1. Introducción

La recuperación de información geográfica (GIR a partir de ahora, del inglés *Geographical Information Retrieval*) pertenece a una rama especializada de la recuperación de información (IR, del inglés *Information Retrieval*) tradicional. Incluye todas las áreas de investigación que tradicionalmente forman el núcleo de la IR, pero además con un énfasis

en la información geográfica y espacial. La recuperación de información geográfica se preocupa de la recuperación de información que involucra algún tipo de percepción espacial. Muchos documentos contienen algún tipo de referencia espacial relevante para la búsqueda (Mandl et al., 2007).

Existen congresos y foros de evaluación como el Text REtrieval Conference¹ (TREC) y el CLEF² que no evalúan expresamente la relevancia en la tarea de la recuperación de información geográfica. El objetivo del Geo-

* Este trabajo ha sido financiado por el Ministerio de Ciencia y Tecnología a través del proyecto TIMOM (TIN2006-15265-C06-03) y el proyecto RFC/PP2006/Id514 financiado por la Universidad de Jaén.

¹<http://trec.nist.gov>

²<http://www.clef-campaign.org>