Online Image Search and Text Categorization Tool

Technology #m07-025

“Lead Inventors: Rebecca J. Passonneau

Online Image Search Challenged by Text Categorization Methods using Word Usage Analysis
Current text categorization methods are based on measurable differences in patterns of word usage between segments. Thus, there is a need for a method to categorize text with regard to features of that text, i.e., images, which cannot be discovered simply through analysis of word distribution. Furthermore, there exists a need for a method for top-down processing of text, determining what words or segments of text would be indicative of a particular categorical label. The invention provides a categorization method using machine learning algorithms and natural language processing to analyze the textual description of an image, extract, categorizes and cross-reference facts about images. The invention provides an efficient solution for booming online image search market.

Scanning Efficiencies with Filter that Inputs Text with Image Identifiers
The invention filters input texts with respect to a list of image identifiers, and generates image id/text association pairs with the text represented as a primary and a secondary layer. The image’s primary layer is a set of sentence- or paragraph-size units to describe the image in the text; the secondary layer uses machine learning mechanism to implement a topical filtering of the primary layer into distinct categories.

Applications:
~ Categorization implement for libraries, museums, media archives and online database, especially image database
~ Categorization implement for blog posts/profiles management on social networking websites

Advantages:
~ Improves the efficiency of the scanning subsequent portions of the text
~ Captures information about a collection in a large database using natural language instead of tedious structured data entry
~ Cross-references and connections can be formed implicitly by the software, without manual entry
~ The back-end structure and the categorization algorithms can be changed without having to re-enter or modify the existing text

Publications:


Licensing Status: Available for Licensing and Sponsored Research Support

Inventors

Rebecca J. Passonneau