Chapter 9

Intelligent Semantic Search Engines for Opinion and Sentiment Mining

Mona Sleem-Amer
Pertimm, France

Ivan Bigorgne
Lutin, France

Stéphanie Brizard
Arisem, France

Leeley Daio Pires Dos Santos
EDF, France

Yacine El Bouhairi
Thales, France

Bénédicte Goujon
Thales, France

Stéphane Lorin
Thales, France

Claude Martineau
LIGM, France

Loïs Rigouste
Pertimm, France

Lidia Varga
LIGM, France

ABSTRACT

Over the last years, research and industry players have become increasingly interested in analyzing opinions and sentiments expressed on the social media web for product marketing and business intelligence. In order to adapt to this need search engines not only have to be able to retrieve lists of documents but to directly access, analyze, and interpret topics and opinions. This article covers an intermediate phase of the ongoing industrial research project ‘DoXa’ aiming at developing a semantic opinion and sentiment mining search engine for the French language. The DoXa search engine enables topic related opinion and sentiment extraction beyond positive and negative polarity using rich linguistic resources. Centering the work on two distinct business use cases, the authors analyze both unstructured Web 2.0 contents (e.g., blogs and forums) and structured questionnaire data sets. The focus is on discovering hidden patterns in the data. To this end, the authors present work in progress on opinion topic relation extraction and visual analytics, linguistic resource construction as well as the combination of OLAP technology with semantic search.

DOI: 10.4018/978-1-4666-0330-1.ch009
INTRODUCTION

As a relatively young sub-field of data mining and computational linguistics, opinion and sentiment mining deals with automatic methods and techniques to extract, analyze and search opinions and sentiments expressed in mostly social media web content (Pang & Lee, 2008).

For companies, knowing what their market thinks and feels about their products and services is vital for their success. Traditionally, this question was addressed by market research using polls and surveys to find out about the customers’ opinion. As subjective content is increasingly available in important amounts on the web, search engines seem to be ideal tools for this task. However, existing tools and methods are yet insufficient for capturing them. In this chapter we present our multi-faceted approach to design a semantic enterprise search engine for opinion and sentiment mining.

As opinion mining is a multi-layered task containing in itself independent areas of research and given the fact that this chapter is limited in terms of space and scope, we do not intend to cover all possible aspects of the subject. We made the decision to leave out certain aspects even though they are currently researched actively as an important part of the project (e.g., opinion mining using statistical methods like supervised or unsupervised machine learning) and to concentrate only selected aspects of our opinion search system, including:

• **Relating opinions and topics** – efficient opinion mining does not only extract opinion polarity (e.g., positive/negative) but also topics (e.g., solar cells) or topic features (e.g., the price of solar cells) and the relations between them (e.g., positive opinions about solar cells). While opinion and topic extraction is well covered in literature, works about relating opinions and topics are still sparse. This chapter presents a method of relation extraction and describes the problems encountered.

• **Extracting opinion and topics / building ontologies and dictionaries** – we present a method to create linguistic resources for opinion and topic extraction according to a semantic opinion/sentiment model we developed for the project. Existing applications often limit opinion search to classifying entire web pages or sentences into either positive or negative categories. Our approach is to take into account ambiguity, intensity and negation and refine the binary positive negative scheme to make subjectivity information more relevant for real-life business decision making.

• **Combining OLAP technologies with a search engine** – as search engines usually focus on opinions in individual web documents they often fail to give a measurable overview of the entire document set. However, companies need opinion metrics for benchmarking and decision making. Business intelligence tools like OLAP (online-analytical processing) address this need by enabling multidimensional data storing and benchmark indicator calculation. As these tools could benefit from semantic search engine functionality, we attempt to combine the best of both in one tool.

• **Designing an opinion search user interface** – we have conducted a user needs study for two distinct business use cases and data sets involving companies in the market research (video games) and energy sector. By taking a user centered perspective, we present new strategies for querying, ranking and visualization for opinion search engine interface design. HMI evaluation tests will be conducted in the last phase of the project by the Lutin laboratory. However, these tests are not subject of this chapter.

Related Content

Speech Emotion Analysis of Different Age Groups Using Clustering Techniques
[www.igi-global.com/article/speech-emotion-analysis-of-different-age-groups-using-clustering-techniques/193250?camid=4v1a](www.igi-global.com/article/speech-emotion-analysis-of-different-age-groups-using-clustering-techniques/193250?camid=4v1a)

Indexing the World Wide Web: The Journey So Far
[www.igi-global.com/chapter/indexing-world-wide-web/64418?camid=4v1a](www.igi-global.com/chapter/indexing-world-wide-web/64418?camid=4v1a)

Question Answering
[www.igi-global.com/chapter/question-answering/64431?camid=4v1a](www.igi-global.com/chapter/question-answering/64431?camid=4v1a)

Latent Topic Model for Indexing Arabic Documents
[www.igi-global.com/article/latent-topic-model-for-indexing-arabic-documents/126329?camid=4v1a](www.igi-global.com/article/latent-topic-model-for-indexing-arabic-documents/126329?camid=4v1a)