Chapter 17
On the Relationship between Ontology-Based and Holistic Representations in a Knowledge Management System

Sylvia Melzer
Hamburg University of Technology, Germany

ABSTRACT
This chapter presents a way for systematically combining ontology-based and holistic-based content descriptions in the context of knowledge management in order to increase recall while maintaining precision.

INTRODUCTION
Usually, Knowledge Management (KM) systems are based on Content Management (CM) systems. In CM systems, content is stored, organized, and supplemented by metadata. Among simple descriptions for authors, characters, publishers, and so on, nowadays, metadata contains feature-based as well as ontology-based content descriptions (see Figure 1). For representing ontology-based content descriptions formal languages are used (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003). For instance, ontology-based content descriptions can be represented via logic-based techniques (Kaya, 2011; Espinosa, 2011).

Applications exploit ontology-based content descriptions in various ways, for example, in the Semantic Web content descriptions are used for finding documents, images, videos, or persons. Search requests are specified by posing queries.
On the Relationship between Ontology-Based and Holistic Representations in formal languages such as string patterns, logic-based queries, and so on. For matching queries with context, each query language has its pros and cons (Melzer, 2006). For most purposes, string patterns have a high recall but do not lead to high precision. Recall is defined as the number of relevant items retrieved divided by the number of relevant items in the repository. Precision is defined as the number of relevant items retrieved divided by the overall number of retrieved items. In practice it is difficult to maximize precision and recall simultaneously.

Until now, Information Retrieval (IR) processes were rarely based on ontology-based content descriptions for matching queries with content. Usually, so called holistic content descriptions and corresponding similarity measures are used for information retrieval (Manning, Raghavan, & Schütze, 2008). Matches on holistic content descriptions can be realized, for example, using nearest-neighbor algorithms (Manning, et al., 2008, pp. 403–419).

In short and slightly exaggerating, holistic representations lead to high recall and low precision, and ontology-based representations tend to be characterized by low recall and high precision (Blank, Meeden, & Marshall, 1992; Espinosa, Kaya, Melzer, Möller, & Wessel, 2007a; Espinosa, et al., 2007b). In this context, it is desirable to increase recall while at least maintaining precision. This kind of improvement could be achieved by systematically combining ontology-based and holistic content descriptions. To the best of our knowledge, a combination of both kinds of content descriptions has not yet been investigated in a methodological way. It is a central idea of this chapter to suggest a way for systematically combining ontology-based and holistic content descriptions in order to increase recall while at least maintaining precision. In the following, retrieved documents with high precision will be called high-quality documents.

Retrieval of high-quality documents is a frequent task in KM contexts, in the sense that the documents themselves or, in some applications, their authors are subjects of further steps in KM processes. However, finding documents might be a problem in case that there is no direct match with simple queries. Consequently, queries need to reformulated, which usually is a rather difficult task for users. This is true for pattern-based as well as logic-based queries (Melzer, 2006). Indeed, if there are at least some query results, we argue that these results can be analyzed and exploited...