Perceptual Perspectives for Experience Items: Representation and Query Processing

Christoph Lofi
Manuel Valle Torre

Web Information Systems Group
Delft University of Technology
Delft, Netherlands

lofi@tudelft.nl

Social judgements like comments, reviews, discussions, or ratings have become a ubiquitous component of most Web applications, especially in the e-commerce domain. Now, a central challenge is using these judgements to improve the user experience by offering new query paradigms or better data analytics. Recommender systems have already demonstrated how ratings can be effectively used towards that end, allowing users to semantically explore even large item databases. In this paper, we will discuss how to use unstructured reviews to build a structured semantic representation of such items, enabling the implementation of user-driven queries. Thus, we address one of the central challenges of Big Data: making sense of huge collections of unstructured user feedback.

More specifically, we discuss the challenge of building structured, but latent representations of "experience items" stored in a database (like movies, books, music, games, but also restaurants or hotels) from unstructured user feedback. Such representations should encode the consensual perception of an item from the perspective of a large general user base. If this challenge could be solved, established database techniques like SQL-queries, query-by-example, similarity queries, but also several data mining techniques like clustering could be easily applied to user-generated feedback. In the following, we will use movies as a use case. However, the described techniques can easily be transferred to any other domain which has user ratings or reviews available.

While there have been previous works also aiming at representing items in a database based on social judgements (e.g., [1], [2]), we address one major yet unresolved problem: user judgements are inherently subjective as they represent a user's *perception*, and merging all individual judgements into one consensual view can be problematic: this approach works sufficiently well for rating-based systems commonly employed in recommender systems as e.g., [3] which rely on factorizing the user-rating-item matrix, as ratings do not allow to express rich opinions and are very numerous allowing to uncover patterns within user groups. However, semantic quality quickly deteriorates when richer sources as for example reviews are considered ([1] relying on document embeddings for reviews): they are less numerous, but much richer in content and demand the use of sophisticated text-mining techniques like aspect-oriented sentiment analysis [4] or document-embeddings [1] to create structured representations which can be used for database query processing. Combining the resulting representations (as for example by averaging the document embedding vectors, or aspect-sentiment tuples) into a single tuple is often not meaningful: Consider for example the movie "Twilight (2008)", typical reviews might express that this is a "beautiful romance full of alluring characters" or "an overall stupid movie and a disgrace to the vampire genre". Some reviews might even be unrelated to the item itself, e.g., "horrible service from vendor, my DVD was damaged on arrival" – averaging the representation of these opinions is mostly meaningless.

Therefore, in this paper, we propose to represent each experience item in a database using *multiple shared perceptual perspectives*, with each perspective representing one major consensual opinion aggregated from multiple homogenous user judgements. In this regard, our contributions are:

- We present the theoretical foundations of shared perspectives in the context of relational databases
- We provide an overview of the design space of different techniques and methods available to obtain and process such perspectives
- We introduce an adapted variant of the query-by-example paradigm intended to interact with multiple shared perspectives
- We present our prototype implementation, and give insights into its query processing performance using simulations on real-life data
- We conduct and present a user study, giving insights into the usefulness and semantic representativeness of the perspectives in our prototype system

REFERENCES

[1] C. Lofi and P. Wille, "Exploiting social judgements in big data analytics," in *13th Lernen Wissen Adaption (LWA) Conference*, 2015, vol. 1458, pp. 444–455.

[2] C. Lofi and C. Nieke, "Exploiting Perceptual Similarity: Privacy-Preserving Cooperative Query Personalization," in *Int. Conf. on Web Information System Engineering (WISE)*, 2014.

[3] J. Selke, C. Lofi, and W.-T. Balke, "Pushing the Boundaries of Crowd-Enabled Databases with Query-Driven Schema Expansion," *Int. Conf. Very Large Data Bases*, vol. 5, no. 2, pp. 538–549, 2012.

[4] M. Ye, C. Lofi, and N. Tintarev, "Memorability of Semantically Grouped Online Reviews," in *Semantics 2017*, 2017.