

# Towards Human-Centered Database Query Processing using on Perceptual Properties

Christoph Lofi

Web Information Systems Group  
Delft University of Technology  
Delft, Netherlands

lofi@tudelft.nl

Some of the most valuable features of Relational Databases are clearly defined schemas with crisp semantics, thus allowing for rich and complex declarative queries. However, this also comes at a cost: the underlying schema must be carefully designed upfront in order to support the queries expected to fulfill the information need of the users of the intended application, and the modelling of the structured schema should represent the actual nature and semantics of the represented real world entities in such a way that it naturally aligns with the internalized semantics of user issuing the queries. Here, in some application scenarios, this focus on strict schemas can become problematic. As an example, consider an e-commerce scenario focusing on selling *experience* products like movies, books, music, or games. Here, the *perceived properties* describing the user experience those products will entail (which, for most people, is the deciding factor for buying the product) are difficult to capture using relational schemas, which thus often leads to a focus on more objective and crisp properties like production year, actor names, or rough genre labels. As a result, many queries users would naturally ask are not supported by the system, as for example queries for movies which “feel” like a given example movie, or movies which feature a “thought-provoking plot”, movies which are “educational”, or “suitable for children” (we call those queries *human-centered queries*, as they are the queries most humans would use in a natural conversation with another human, but are often not supported by information systems.) One of the challenges around perceived properties of experience products is that it is very hard to foresee during schema design time which properties will be relevant for users, and how they are perceived by them (i.e., the challenge of actually obtaining values for the properties.) Especially, many of these properties might even be subjective, and thus the perception of different users might differ or be even conflicting (e.g., there might be conflicting views on how “funny” a given movie is).

We claim that most of the perceptual information required to actually support such human-centered queries can be obtained from user *user-generated judgements* as for example ratings, comments, or reviews, as such judgements usually cover the perceptual properties and aspects deemed important by the creator of the judgement. Thus, such judgments can be very valuable in supporting a user’s decision process when the natural query capabilities of the

underlying database system are lacking. However, integrating this rich source of information into the query process is hard due to the aforementioned challenges, and many applications choose not to try an integration at all: e.g., in most applications (like for example web shops), user reviews are simply displayed for manual consultations, or user ratings might be used within a recommender systems – but usually it is not possible to access the richness of information contained in human judgements in a declarative and explicit relational fashion.

In this work, we are discussing the challenge of supporting such human-centered queries focusing on perceptual properties from a database query processing perspective.

Our contributions are as follows:

- We present a general vision of a database system using perceptual properties, and discuss a high-level model of how to integrate *perceptual properties* into the relational model
- As a part of that vision, we introduce and discuss the concept of *consensual perceptual* properties to deal with subjectivity in user perception, i.e., properties of entities for which the values emerge form a consensus in perception of a larger user base. Also, we introduce *multi-consensual properties* for which there is not a single, but multiple consensual values.
- We further classify perceptual properties into *explicit* and *latent* properties, and discuss the role of this classification in our model. Here, explicit properties have a real world interpretation which is quickly obvious to users, while latent properties are opaque but still can be used for several query types like similarity queries.
- We discuss the *high-level impact* of these design choices on database schema modelling process, and on database query processing.
- We present and discuss the results and experiences gathered from our prototypical implementation of such a system which focuses on *consensual latent properties* of movies, and which uses Amazon reviews as a source for user-generated judgements.
- We outline how our prototype system could be expanded to cover the full semantics of our vision.