



# Subjective Database

Sivaporn Homvanish, Tzu-Man Wu  
UPC Universitat Politècnica de Catalunya, Big Data Seminar, eBISS 2019



## Abstract

Nowadays, more and more people rely on online search and providing the results that fit user's need becomes highly important. Most of the traditional databases support only objective query. However, because the online experiential search are often subjective and imprecise, subjective query and subjective database appeared to be a possible solution to return more customized results. There are many researches developed the processes and algorithms on extracting subjective expressions and opinions, however, challenges to combine subjective query and subjective data still exist. In this poster, it provides a summary of data model, query processing, subjective database implementation from OpineDB[1] which managed subjectivity on both data and queries ends and used a schema of linguistic domains and marker summaries. Moreover, this report introduced the used of subjective database with various applied fields and its challenges.

## Introduction

### Subjective Data and Query

Subjective means the information which based on or influenced by personal feelings, tastes, or opinions. Both subjective data and query are different depending on personal thoughts. On the contrary, objective data and query are measurable and have the same interpretation among different people.

Query	subjective	Flights to London with reasonable price	Airlines that serve delicious food
	objective	List all flights to London <=200€	Airlines with average food rating >4.9
		objective	subjective

Figure 1: Subjective and objective queries/data

## Subjective database OPINEDB Methods

### Data Model in Subjective Database

- Linguistic Domain(LD): It is a set of phrases that describes the attribute.  
e.g. Satisfaction LD { "very satisfied", "satisfied", "average", "unsatisfied", "very unsatisfied", "not satisfied", "very bad", "dislike", "disappoint", "comfortable", "happy" ...}
- Marker Summary: It represents the important concept of the domain.
  - Linearly-ordered: markers have linear-scale form  
e.g. In-flight services satisfaction ["satisfied", "average", "unsatisfied"].
  - Categorical: markers are categorized into several types  
e.g. Customer on the flight["first class", "business class", "economy class"]

### Processing subjective database queries



Figure 2: Subjective queries against subjective data

There are 3 steps to process the query

- Step 1: Predicate Interpretation  
This step is to match subjective attributes from the database with the value that user is specifying.
  - Word embedding method
  - Co-occurrence method
  - Text-retrieval method
- Step 2: Compute degrees of truth  
This method is used to indicate how well of the marker summary represented the query predicate.
- Step 3: Fuzzy logic  
Combine the degrees of truth and return results.

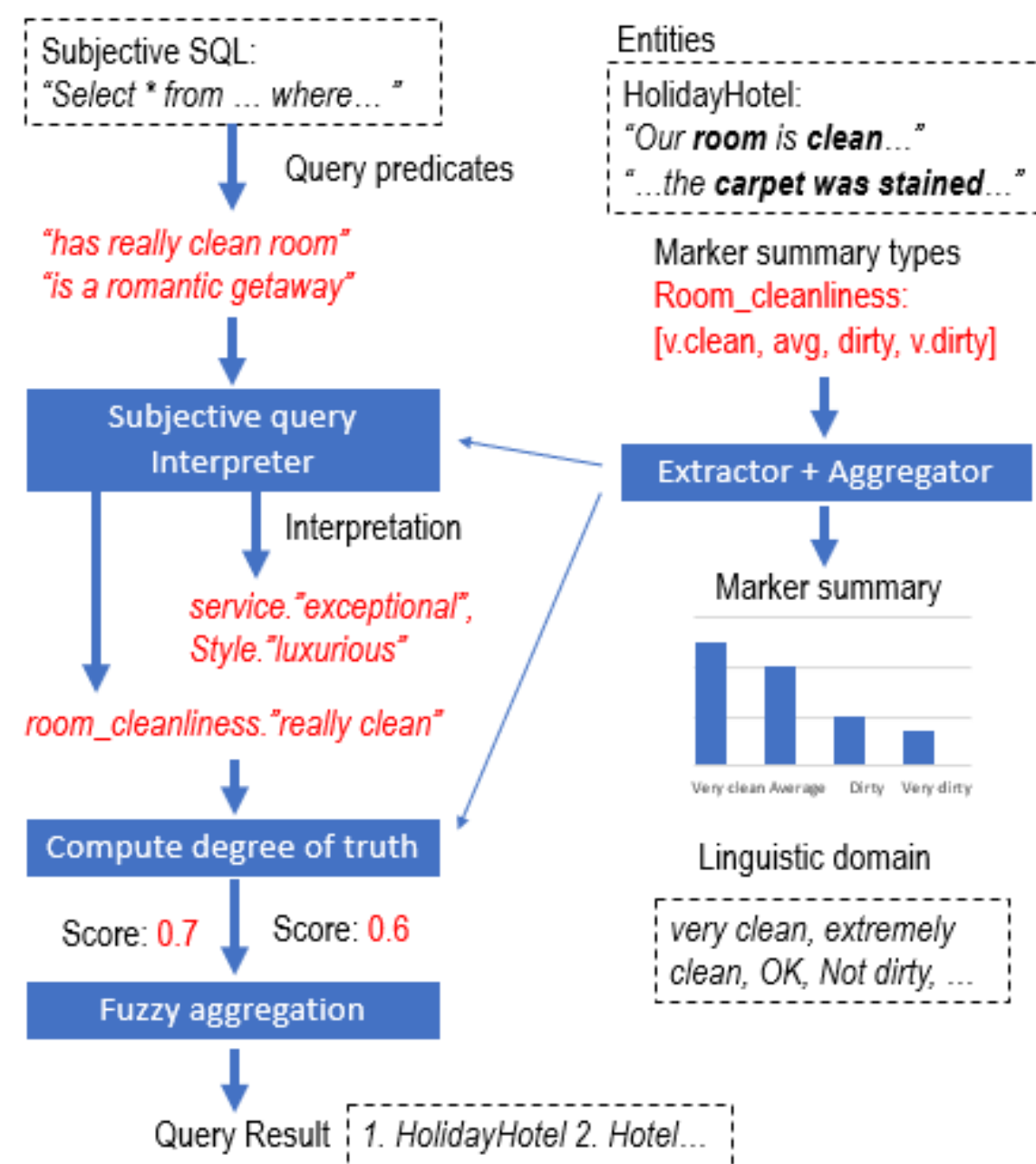


Figure 4: Processing subjective queries with OPINEDB

## Designing subjective database

After discussing how to process subjective query. Next, the following steps describe how subjective data is created in OpineDB.

### Step1: Extracting opinions from reviews

The room was very clean, but the staff was not so friendly.

extract pairs (aspect + opinion)

{{"room", "very clean"}, {"staff", "not so friendly"}}

### Step2: Design schema for subjective attribute

room\_cleanliness, staff\_attitude, .....

### Step3: Map linguistic variation to subjective attribute

Text classification ("room", "very clean") → Attribute A : room\_cleanliness

generate training set for high quality classifier

Seed expansion E = {"room", "bedroom", "carpet", "furniture"} + {"suite", "apartment"}

P = {"clean", "dirty", "very clean", "very dirty"} + {"stained", "dusty"}

OpineDB constructs a labeled tuple (concat(e, p), A)

where concat(e, p) is the example and the attribute A is the label.

test

235 seeds of 15 attributes → 5000 tuples → 88% accuracy

### Step4: Define markers

Linearly-ordered "dirty" < "average" < "clean" < "very clean" sentiment analysis

Categorical bathroom attribute {"luxurious", "modern", "old-styled"} k-means clustering

### Step5: Compute marker summaries

OpineDB Aggregation Function: count the number of phrases in the reviews

ABC Hotel room\_cleanliness : [ very\_clean: 90.5, average:60.5, dirty: 30, very\_dirty: 20 ]

Figure 5: Designing subjective database

## Applied Fields

This section explains more about the data that can be implemented with subjective database.

- Text Mining: There are many researches about text sentiment analysis and opinion mining. OpineDB also creates database schema by using customer reviews of the hotel which has text format as a source data.
- Speech Recognition: A research[6] developed three top-level goals which is objective, subjective, and navigation for voice search recommendation. Then, it applied multiple goals in a single request. For example, "drama and good movies" includes one objective goal and one subjective goal ("drama": genre, "good movies": subjective goal). Then, the system will count the frequency of different recommendation goals inside their database.
- Image and Video Processing: By giving example images, system determines the relative preferences by extracting image feature and store it in appropriate representation. When user query, system retrieves data in the database to find user preference and combine with user query to display ranked image as a result [7]

## Challenges

- No standard schema to follow
- Apply suitable data aggregation function
- Difficult to support the variety of linguistic phrases in the specific domain
- No attribute suggestion for schema designer
- Does not consider user profile and preference
- Hard to detect some specific emotion of users

## Conclusion

This report first introduced OpineDB[1] which managed subjectivity on both data end and queries end and also used a schema of linguistic domains and marker summaries. Then, it provided three applied fields which are text mining, speech recognition and image/video processing related to subjective data. However, there are still some missing parts mentioned in Challenge can be improved. In the future, the development of subjective database will be more complete if also considering user's preference and specific emotion, applying the most suitable aggregation function and attribute suggestion mechanism for designer and building a standard schema if possible.

## Reference

- [1] Yuliang Li, Aaron Feng, Jinfeng Li, Saran Mumick, Alon Halevy, Vivian Li, Wang-Chiew Tan. Subjective Databases. arXiv preprint arXiv:1902.09661, 2019
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [4] Text Learning Group. (2019). Retrieved from <http://sentiwordnet.isti.cnr.it/>
- [5] Chinsha, T.C. & Joseph, S. (2014) Aspect Based Opinion Mining from Restaurant Reviews. International Journal of Computer Applications. Aspect based Opinion Mining from Restaurant Reviews
- [6] Jie Kang, Kyle Condiff, Shuo Chang, Joseph A. Konstan, Loren Terveen, F. Maxwell Harper. Understanding How People Use Natural Language to Ask for Recommendations(2017). <http://files.grouplens.org/papers/harper-recsys2017.pdf>
- [7] Wei Zhu; Rajiv Mehrotra(2002).United States Patent No.US 6,345,274 B1 Retrieved from <https://patents.google.com/patent/US6345274B1/en>