

Jannik Strötgen\*, Omar Alonso<sup>+</sup>, Michael Gertz\*

\*Database Systems Research Group, Heidelberg University, Heidelberg, Germany

<sup>+</sup>Microsoft Corp., Mountain View, California, USA

## Motivation

### Reviews are valuable

- for customers → purchase decision
- for manufacturers → collect opinions

### Typical Exploration Functionality

- total number of reviews
- average rating
- number of 1, 2, 3, 4, 5 star ratings
- sort by helpful votes or recency

→ no analysis of all reviews collectively  
→ no temporal analysis

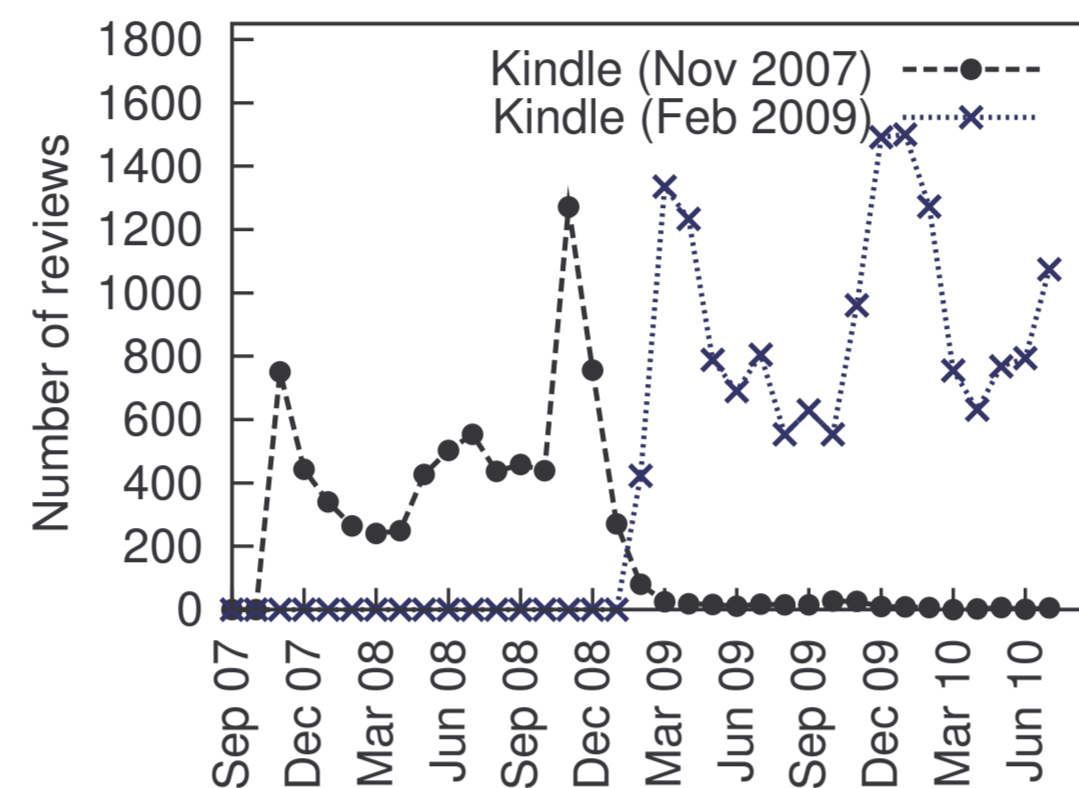
### Review sets of products

- different in size
- different in rating

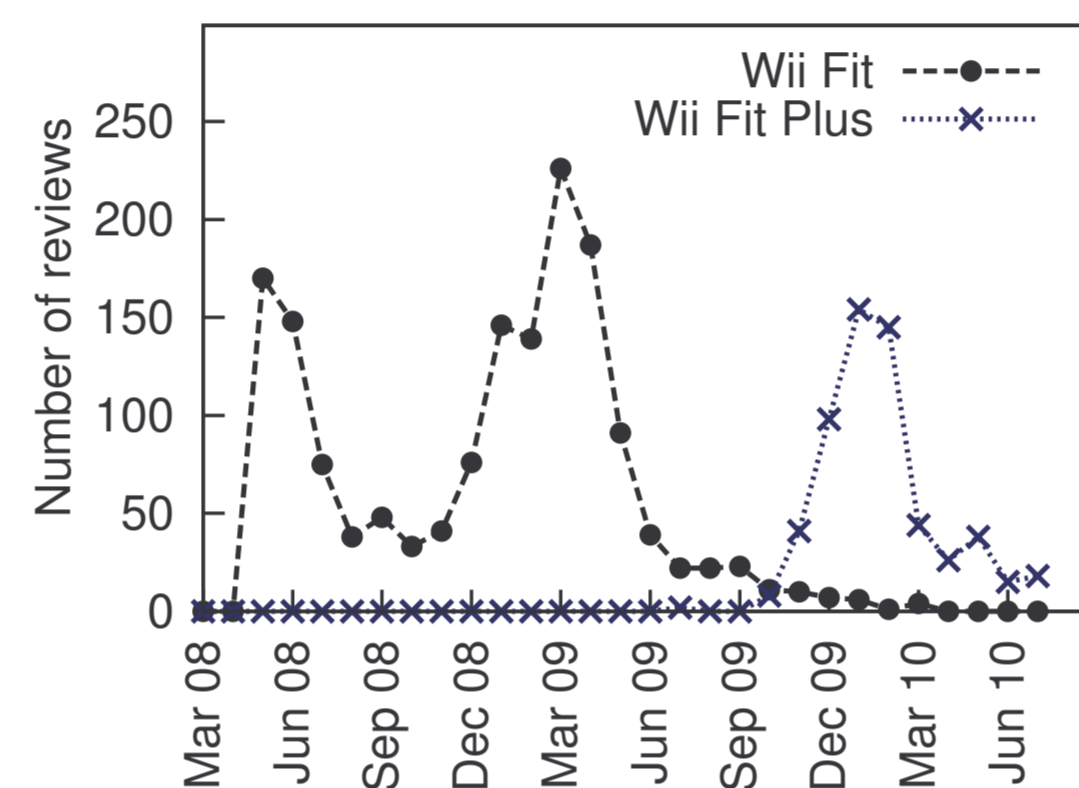
### Review sets seem to be similar...

→ all products with avg rating > 4,0  
... but are different  
→ different temporal characteristics

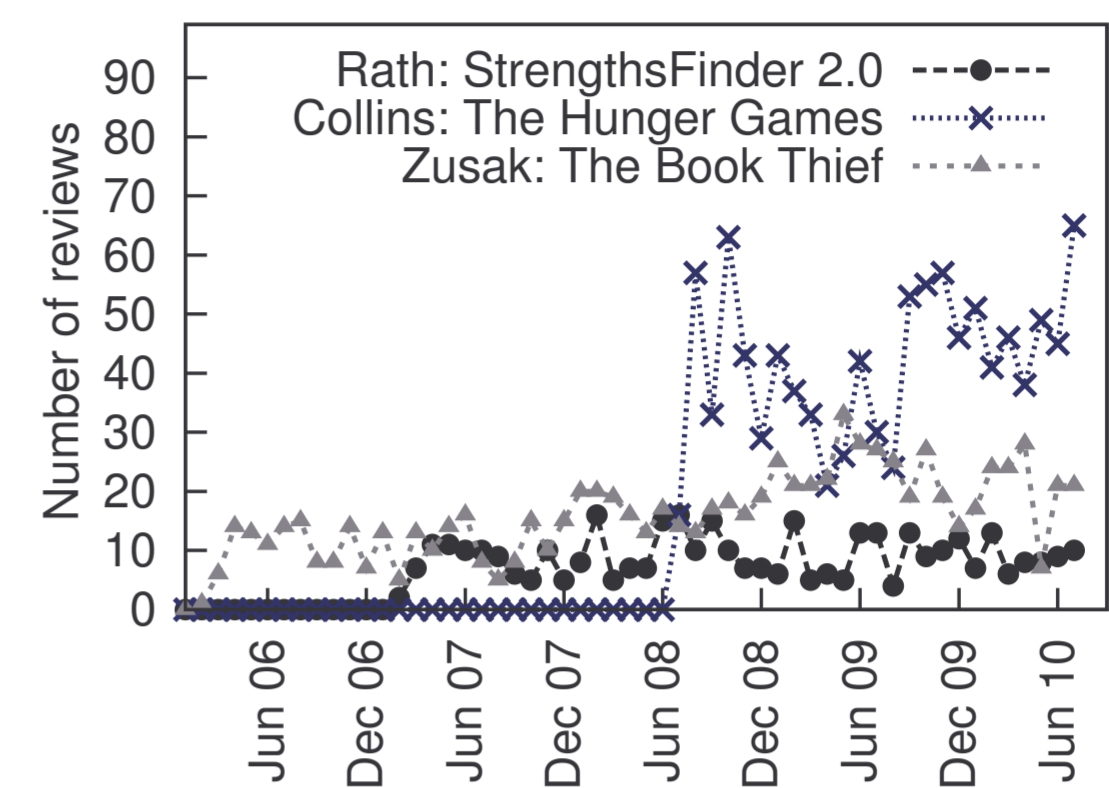
### a) Substitutes I



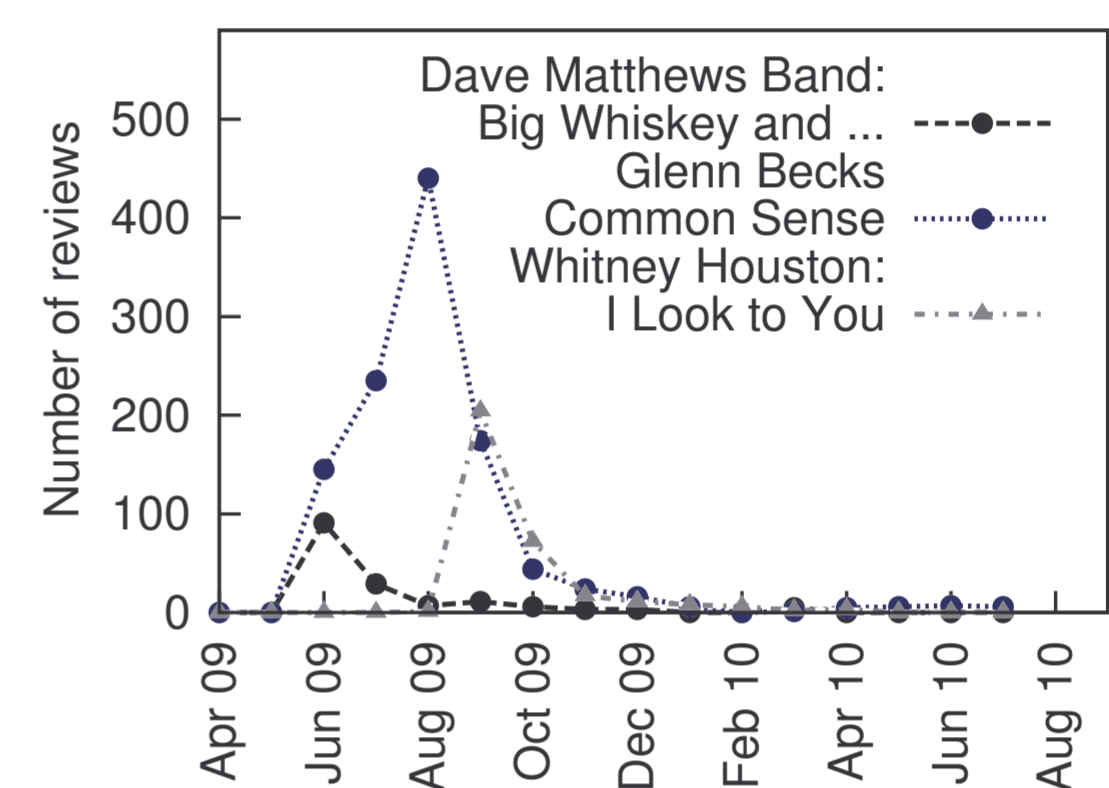
### b) Substitutes II



### c) Evergreens



### d) Bursts



## Retro: Key Features

### Temporal Cluster Constructor

- timeline-based clustering
- buckets with decreasing granularity

### Sparkline Constructor

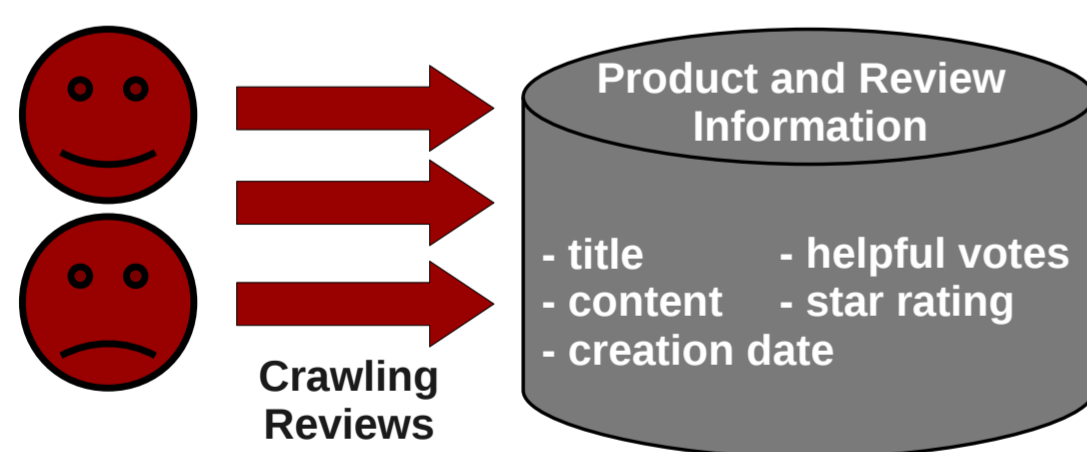
- temporal distribution of reviews
- star rating distribution in each cluster

### Balanced Review Selector

- the high-rated and low-rated reviews with most helpful votes in each cluster

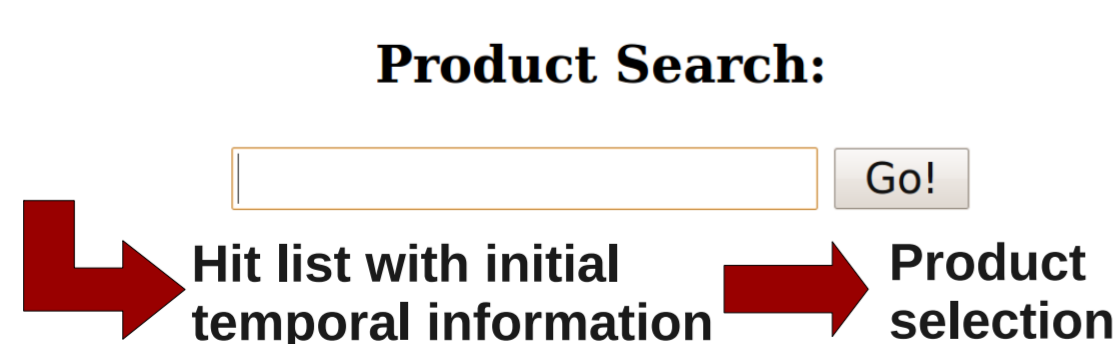
## Retro: Architecture

### 1) Preprocessing Phase

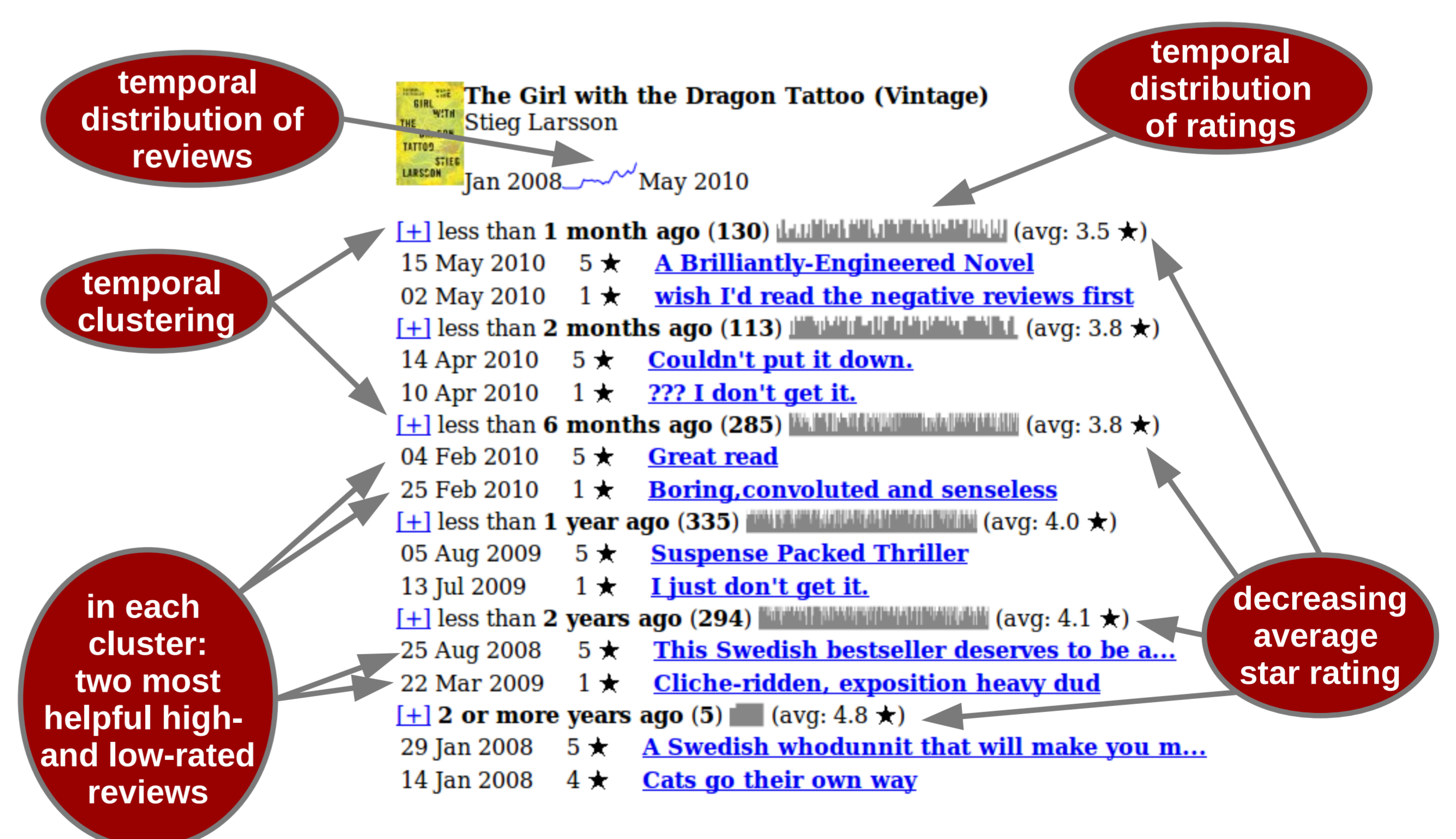


### 2) Search Phase

#### Retro - Temporal Exploration of Reviews



## Retro: Temporal Exploration of Product Reviews



## Contact information

Jannik Strötgen  
stroetgen@uni-hd.de  
<http://dbs.ifi.uni-heidelberg.de/>