

Special Lecture 12

User in the Loop

Relevance Feedback

User Feedback

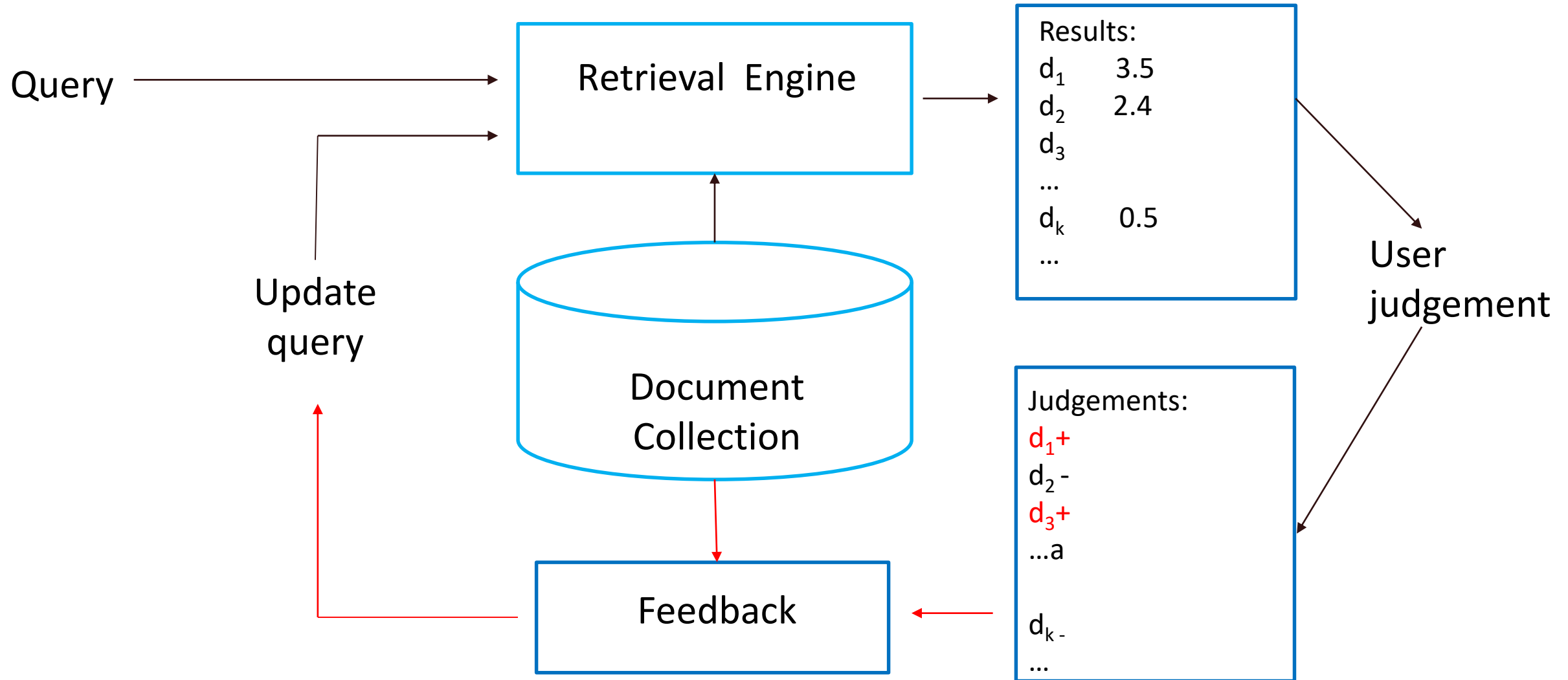
- An IR system could be an interactive system



Use Scenario

- A query q or a classifier w is given
- Search engine retrieves a set of possible answers
 - x_1, x_2, x_3 , etc.
- System guess the user intend and improve the answers
 - x_7, x_{12}, x_{23} , etc.
- User is able to smartly navigate and get what she is looking for.
- E.g. Search for a specific fashion/design in a large database

Relevance Feedback



Effective and Popular(?)

[Personalization - Wikipedia, the free encyclopedia](#) [f] [x]

Personalization involves using technology to accommodate the differences between individuals. Once confined mainly to the Web, it is increasingly becoming a ...
[en.wikipedia.org/wiki/Personalized](#) - 42k - [Cached](#) - [Similar pages](#) - [i]

[Personalized Gifts from Personalization Mall](#) [f] [x]

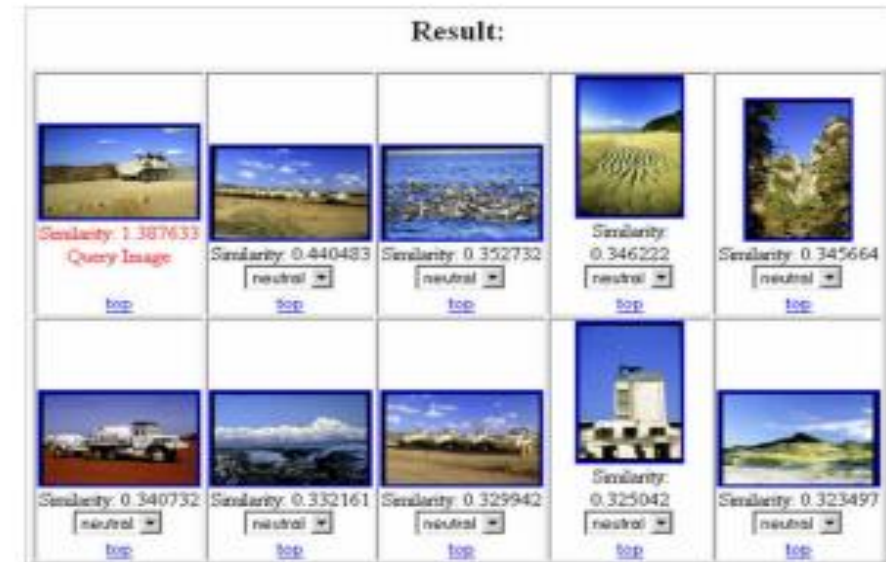
It shows you went out of your way to find the perfect gift at [i] to **personalize** it to make it theirs alone! At PersonalizationMall.com, we design most of our ...
[www.personalizationmall.com/Default.aspx?&did=111028](#) - 47k - [Cached](#) - [Similar pages](#) - [i]

[What is personalization? - a definition from Whatis.com](#) [f] [x]

Mar 6, 2007 ... On a Web site, **personalization** is the process of tailoring pages to individual users' characteristics or preferences.
[searchcm.techtarget.com/sDefinition0,,sid11_gc532341,00.html](#) - 72k - [Cached](#) - [Similar pages](#) - [i]

Relevant

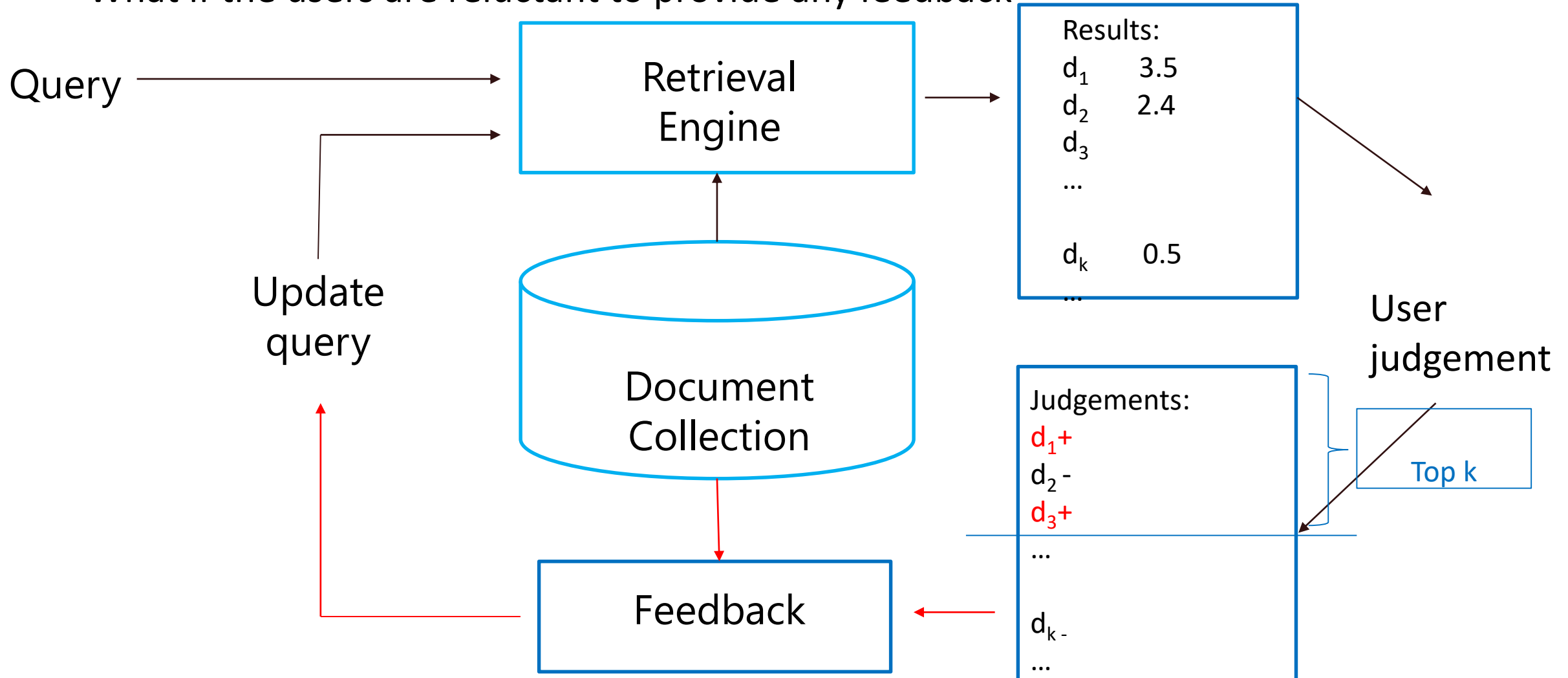
Nonrelevant



Too Explicit?

Pseudo Feedback and Query Expansion

- What if the users are reluctant to provide any feedback



Rocchio Model

$$Q_1 = \alpha Q_0 + \frac{\beta}{n_1} \sum_{i=1}^{n_1} R_i - \frac{\gamma}{n_2} \sum_{i=1}^{n_2} S_i$$

where

Q_0 = the vector for the initial query

R_i = the vector for the relevant document i

S_i = the vector for the non-relevant documents i

n_1 = the number of relevant documents chosen

n_2 = the number of non-relevant documents chosen

α , β and γ tune the importance of relevant and non relevant terms (in some studies best to set β to 0.75 and γ to 0.25)

Illustration (NN View)

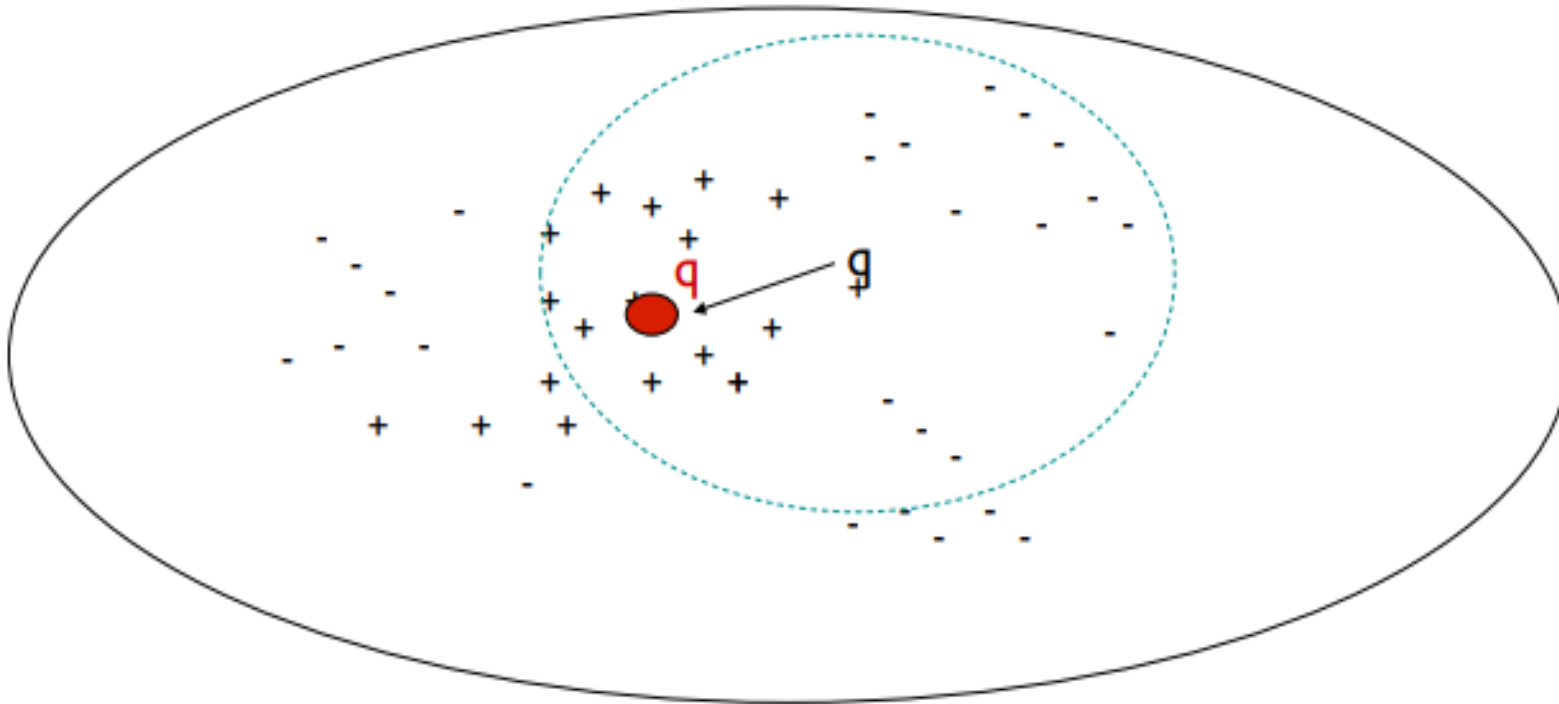
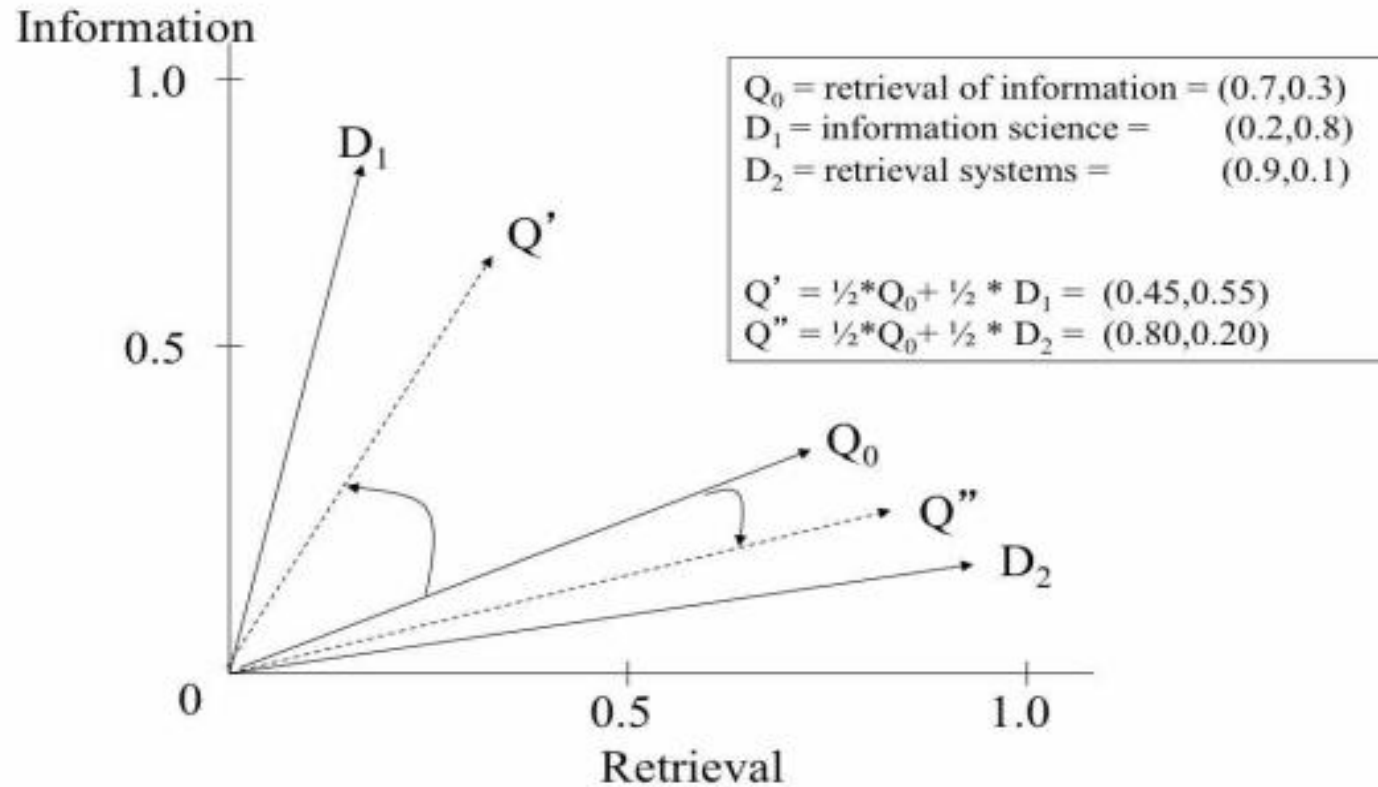


Illustration (dot product view)



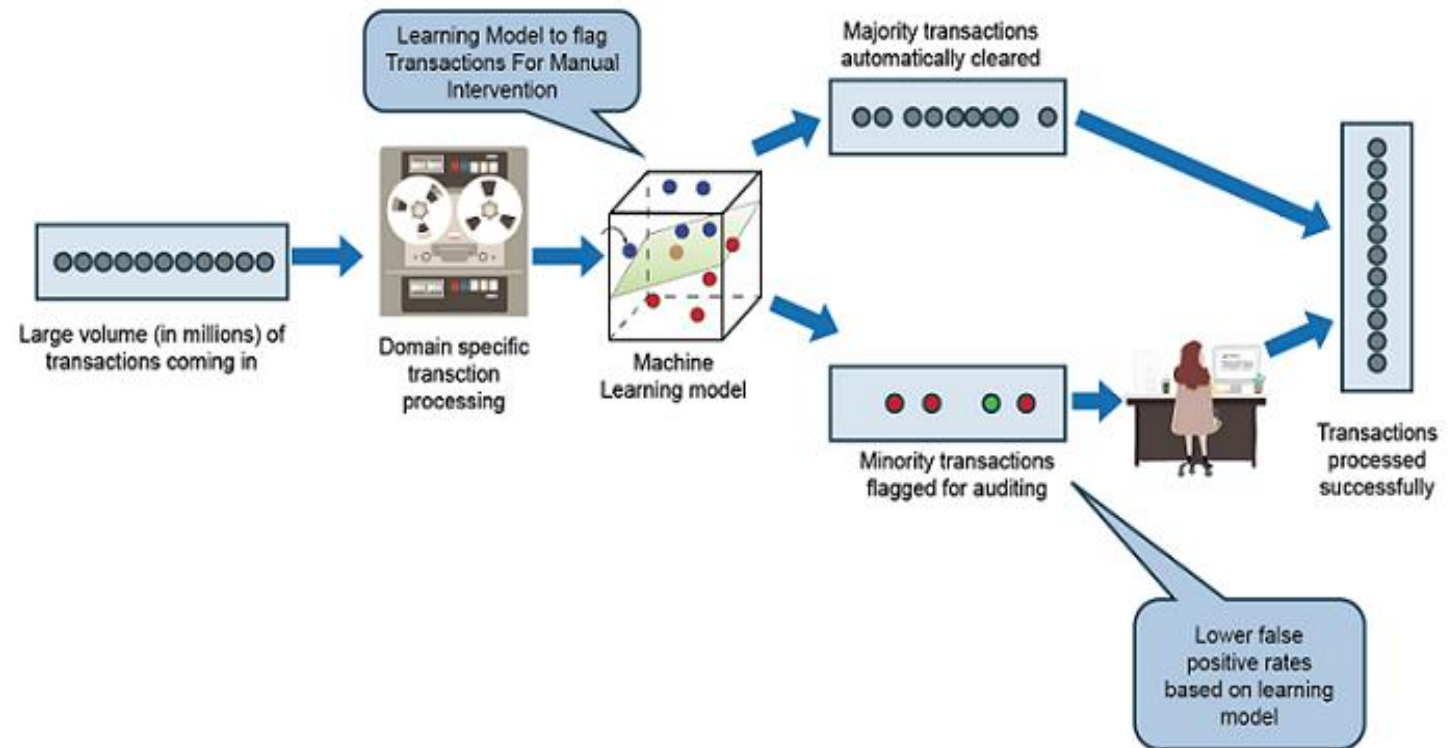
Challenges and Refinements

- Can we force user to say + and – on the answers?
 - Often + is more clear ?. But not – ve is not shared.
 - Cases when only + or Only – is available.
- Often + is implicit (I click/browse) and not explicit.
- Examples:
 - Browsing for fashion (clothes)

Users in the ML systems

Interactive Classification

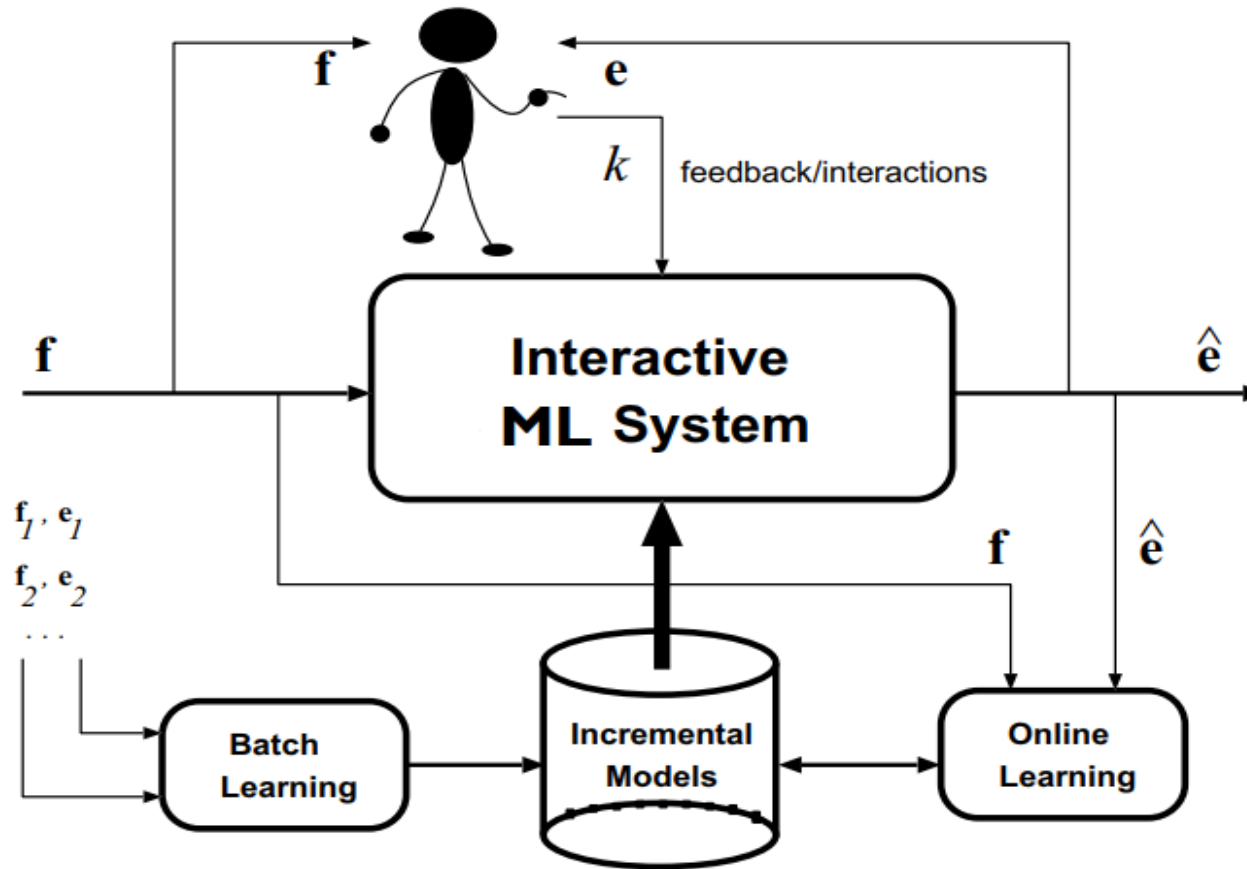
Goal: Optimize life-time Return On investment



Defining Characteristics

- Expensive domain experts
- Skewed class distribution(minority events)
- Concept/ Feature drift
- Biased sampling of labeled historical data
- Lots of unlabeled data

Another Scenario (Interactive ML)



User may adjust:
- K (control) or
- Directly give desired e

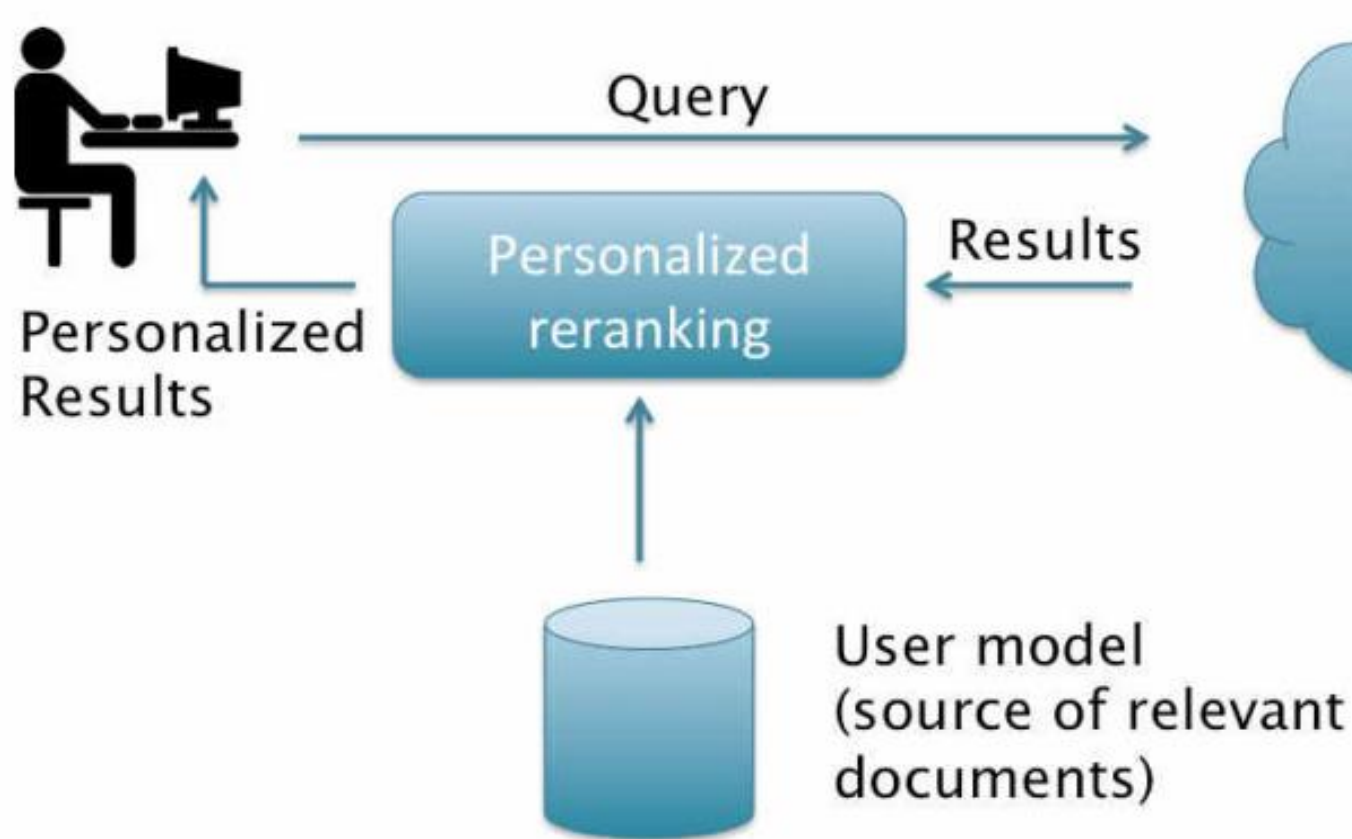
Associated Issues

- Incremental and Computational Issues
 - How do we learn, adapt and forget
 - What is the basic knowledge and what do we adapt?
- Stability
 - Am I overlearning and changing too fast?
 - Stability, convergence and other algorithmic issues.

Personalization



Use Case (RF style)



Many Relevant Attributes

- Consider the following pieces of information
 - Geographical Location
 - Age, gender, ethnicity, religion, etc.
 - Interests
 - Previous reviews on products
 -
- How could these pieces of information help?
- How to collect these information?

Approaches

- Individual Vs Collaborative
- Reactive Vs Proactive
- User Vs Item Information

Individual Vs Collaborative

- Individual approach (E.g. Google Personalized Search)
 - Use only individual user's data
 - Generate user profile by analyzing
 - User's browsing behavior
 - User's active feedback on the system

Individual Vs Collaborative

- Advantage
 - Can be implemented on the client-side - no privacy violation
- Disadvantage
 - Based only on past interactions.

Reactive Vs Proactive

- Reactive approach
 - Explicitly ask user for preferences
 - Either in the form of query or feedback
- Proactive approach
 - Learn user preferences by user behavior
 - No explicit preference demand from the user

Reactive Vs Proactive

- Behavior is extracted
 - Click-through rates
 - Navigational pattern

User Vs Item Information

- User Information
 - Geographic location (from IP address)
 - age, gender, marital status, etc. (explicit query)
 - Lifestyle, etc. (inference from past behavior)
- Item Information
 - Content of Topics – movie genre, etc.
 - Product/ domain ontology

Thanks!!

Questions?