

Основы обработки текстов, лекция 10: Информационный поиск

Ярослав Недумов

13 ноября 2019



С чего бы вообще что-то искать?



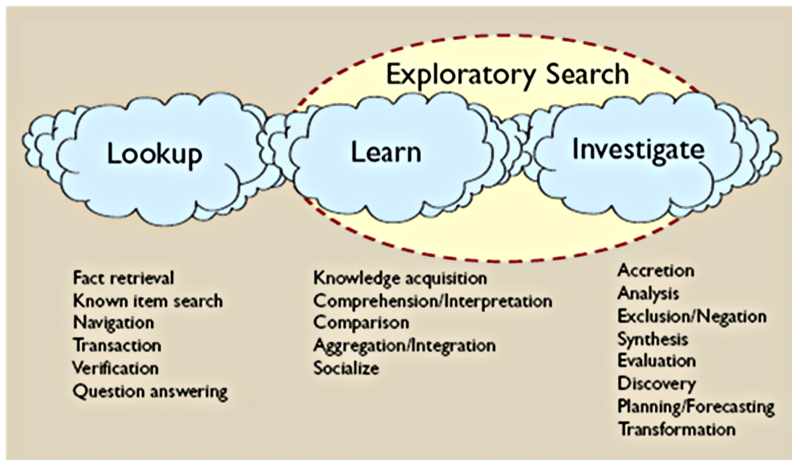
С чего бы вообще что-то искать?



- Неполнота в модели мира
- Любопытство (или необходимость)

- Понять, что же пользователь хочет узнать
- Найти ответ в базе знаний

Степень осознания потребности может быть разной¹



¹Marchionini, Gary. "Exploratory search: from finding to understanding." Communications of the ACM 49, no. 4 (2006): 41-46.

Какие бывают запросы

- Язык запросов
- Ключевые слова
- Вопрос в свободной форме
- Подсказки
- Поиск похожих
- Реклама
- Рекомендации

Какие бывают домены

- Веб
- Фильмы, музыка, картинки
- Люди
- Товары
- Маршруты, карты
- Цифровые библиотеки

- PubMed
- Bing Images
- Навигатор Кинопоиска
- SemanticScholar
- Яндекс Маркет

Совместный поиск: SearchTogether²

The screenshot displays the SearchTogether web interface, a collaborative search tool. At the top, a search bar contains the query "SearchTogether: diabetes (google)". Below the search bar, a list of participants (George, Rachel, Betty) is shown, each with a small profile picture and a list of their search history. The search results are displayed in a central column, showing various links related to "diabetes friendly recipes". On the right side, there is a featured article from Reader's Digest titled "Diabetes-Friendly Holiday Recipes". The interface is annotated with red letters 'a' through 'j' pointing to specific UI elements:

- a**: Points to the search bar.
- b**: Points to the list of participants (George, Rachel, Betty).
- c**: Points to the search results for "diabetes friendly recipes".
- d**: Points to the "Recommendations" section at the bottom.
- e**: Points to the "Standard Search" button.
- f**: Points to the "Split Search" button.
- g**: Points to the "Multi-Engine Search" button.
- h**: Points to the "You can make this for Thanksgiving, Mom!" section.
- i**: Points to the "Back", "Forward", "Comment", and "Recommend" buttons.
- j**: Points to the "Interactive Features" section on the right.

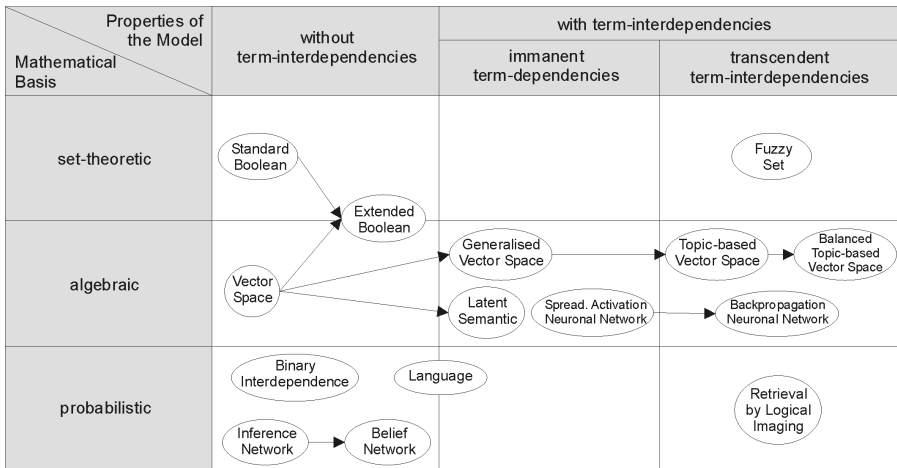
²Morris, Meredith Ringel, and Eric Horvitz. "SearchTogether: an interface for collaborative web search." In Proceedings of the 20th annual ACM symposium on User interface software and technology, pp. 3-12. ACM, 2007.

Классическая модель:

- Ищем текстовые документы
- Боремся с полным перебором
- Боремся за релевантность



Что такое документ?³



³CC BY-SA 3.0

Идем в интернет за картинкой

- Степень совпадения с запросом

- общие слова
- bm25
- ...

- Качество самого документа

- PageRank
- ...

- 10 blue links
- Фасеты
- Кластеры
- + Экзотика

- Явная разметка результатов
- Перехват кликов
- Анализ логов

Снова идем за картинкой в Интернет

