Types of Information Needs
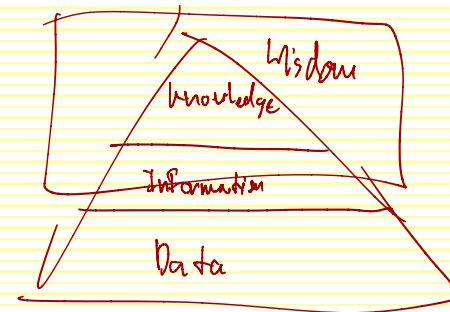
- Retrospective *(Information)*
  - "Searching the past" *ข้อมูล*
  - Different queries posed against a static collection
  - Time invariant

- Prospective
  - "Searching the future"
  - Static query posed against a dynamic collection
  - Time dependent

ปัญหาอะไรได้อันนึงคือ เวลามันจะ ได้ query เก็บ
ผลลัพธ์ถ้าจากจะต่างๆ ไป

Wisdom
Knowledge
Information
Data

What types of information?

- Text (Documents and portions thereof)
- XML and structured documents
- Images
- Audio (sound effects, songs, etc.)
- Video
- Source code
- Applications/Web services

Interesting Examples

- Google image search

  http://images.google.com/

- Google video search

  http://video.google.com/

- Finding naked people (seriously!)

  http://http.cs.berkeley.edu/~daf/people.html

- Query by humming

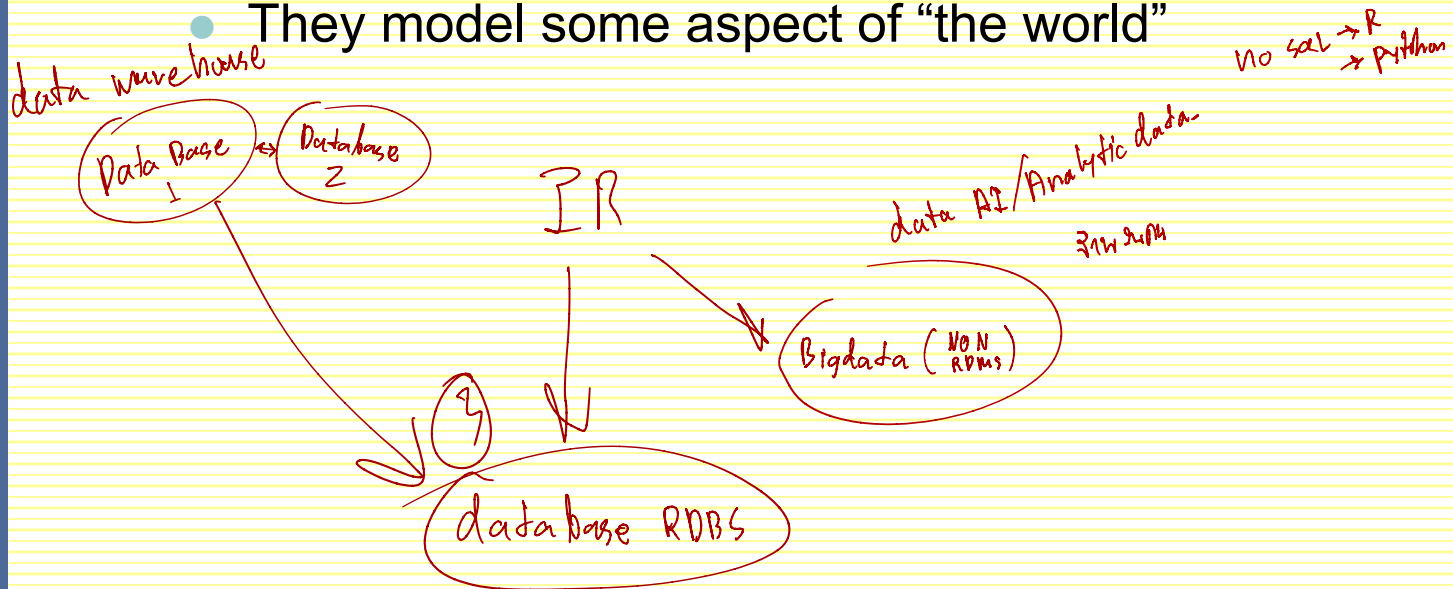 http://www.cs.cornell.edu/Info/Faculty/bsmith/query-by-humming.html

What about databases?

○ What are examples of databases?

- Banks storing account information
- Retailers storing inventories
- Universities storing student grades

○ What exactly is a (relational) database?

- Think of them as a collection of tables
- They model some aspect of "the world"

Data mining

data warehouse

Data Base 1 ↔ *Database 2*

IR

Bigdata (NON RDMS)

no sal → R → Python

data AI/Analytic data

बिग डाटा

database RDBS

Databases vs. IR

(handwritten: gr query Retrospective ... Prospective.)

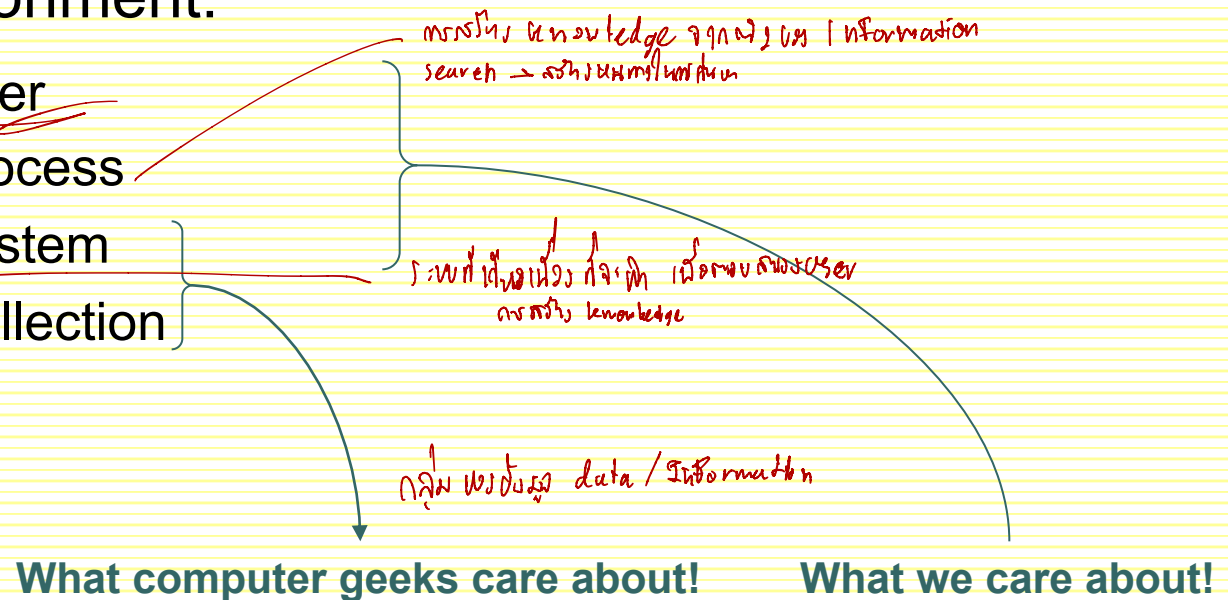|	**Databases**	IR
What we're retrieving	Structured data. Clear semantics based on a formal model.	Mostly unstructured Free text with some metadata.
Queries we're posing	Formally (mathematically) defined queries. Unambiguous.	Vague, imprecise information needs (often expressed in natural language).
Results we get	Exact. Always correct in a formal sense.	Sometimes relevant, often not.
Interaction with system	One-shot queries.	Interaction is important.
Other issues	Concurrency, recovery, atomicity are all critical.	Issues downplayed.

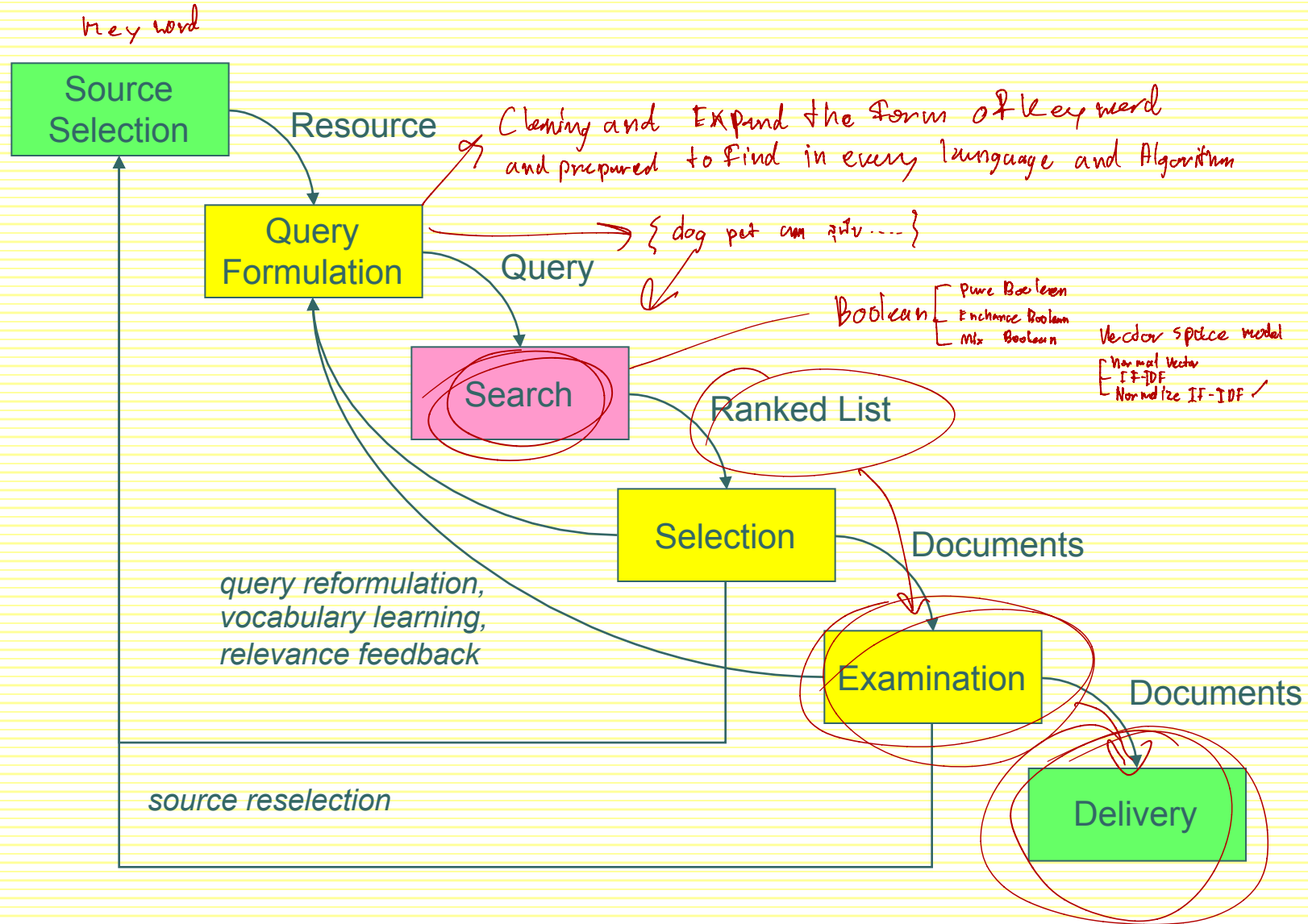# The Big Picture

- The four components of the information retrieval environment:
  - User
  - Process
  - System
  - Collection

การเรียน knowledge จากกลุ่ม User / Information
search → สิ่งที่มนุษย์ในการค้นหา

ระบบที่เกี่ยวเนื่อง ก็จะผลิต เนื้อหาบ ลพธ์User
การสร้าง knowledge

กลุ่ม ทรัพยากร data / Information

What computer geeks care about!   **What we care about!**

The Information Retrieval Cycle

key word

Source Selection

Resource

Cleaning and Expand the form of key word
and prepared to find in every language and Algorithm

Query Formulation

Query

{ dog pet cm बिल्ली .... }

Boolean [Pure Boolean
Enhance Boolean
Mix Boolean]

Vector splice model
[Normal Vector
TF-IDF
Normalize IF-IDF ✓]

Search

Ranked List

Selection

Documents

query reformulation, vocabulary learning, relevance feedback

Examination

Documents

Delivery

source reselection

Supporting the Search Process

Source Selection → Resource → Query Formulation → Query → Search → Ranked List → Selection → Documents → Examination → Documents → Delivery

Indexing → Index → Search

Acquisition → Collection → Indexing

Boolean

Index forms
- document
- website
- database query ...

Simplification?

Source Selection

Resource

Query Formulation

Query

Is this itself a vast simplification?

Search

Ranked List

Selection

Documents

query reformulation, vocabulary learning, relevance feedback

Examination

Documents

source reselection

Delivery