



# Introduction to Information Retrieval

Dra. Mireya Paredes López  
[mireya.paredes@udlap.mx](mailto:mireya.paredes@udlap.mx)

# The evolution of communication



# Activity 1:



1. Discuss with your partner what **Information Retrieval** is about.
2. Write down in your own words, a definition of **Information Retrieval** and circle **keywords**.

# Information Retrieval

**“Information Retrieval”** is a field concerned with the *structure*, *analysis*, *organization*, *storage*, *searching*, and *retrieval* of information.

Source: Search Engines Information Retrieval in Practice, W. Bruce Croft, D. Metzler and Trevor Strohman.



# Digital communication



1. Discuss with your partner what sort of **applications** are used to retrieve information.
2. Give an example of the **structure** of one of them.
3. What is the difference between **text**(book) and a **database record**?. Give an example.

# Structured versus unstructured data

Having the following database register:

***account record, account number and current balance.***

*“Find **account number** 321456”*

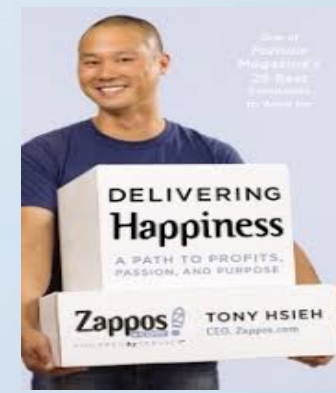
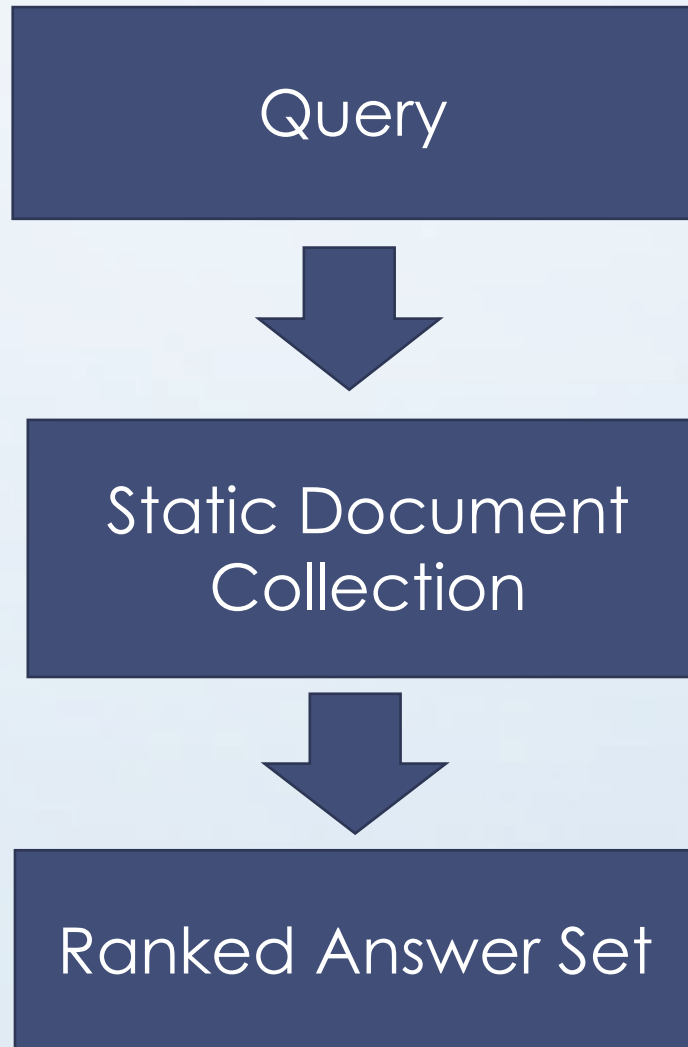
*“Find **accounts** with **balances** greater than \$50,000.00”.*

**What about having a news story?**

*“Meet the former hackers who spend their days stopping the criminals they used to see as their comrades”.*



# Document Retrieval



# Some Dimensions of Information Retrieval

Examples of content	Examples of Applications	Examples of Tasks
Text	Web search	Ad hoc search
Images	Vertical search (particular topic)	Filtering
Video	Enterprise search	Classification
Scanned documents	Desktop search	Question answering
Audio	Peer-to-peer search	
Music		



# The Big Issues in Information Retrieval



1. Relevance
2. Evaluation
3. Information Needs

## Activity 2: 10 minutes



- To read about the Big Issues of Information Retrieval. (pages 4,5,6).
- Make a vocabulary section of the three big issues involved.
- Circle keyword with a red pen.

# Activity 2: 10 minutes



Discuss the following questions with your team mates.

1. Why just comparing the text of a query with the text of a document could get poor results?
2. What refers to a **vocabulary mismatch problem**?
3. What is a solution for the **relevance issue** and why?
4. Why **statistical properties** are an interesting feature of the retrieval models? And give an example.
5. Mention the two metrics for retrieval models evaluation and try to explain them in your words.
6. Mention techniques used to improve *information needs* and *discuss them*. Have you ever used them?