Faculty of SET / School of Computer Science

TECH 1004 AI Technologies
AI Techniques

AI Research

- Learning
  - Example: A weather forecast program that may learn from past weather data to predict the weather in the future.
- Reasoning
  - Example: John is either in the car or in the house. He isn't in the car, therefore he is in the house.
- Problem-solving
  - Example: Finding a winning move in a game.
- Perception
  - Example: A robot that is sensing and then for example cleaning your room.
- Language-understanding
  - Example: A program that answers natural language questions about high-school algebra problems stated in natural language.

Outline

- Data Mining
  - Machine learning
  - Deep learning
  - Pattern recognition
  - Information Retrieval
- Language Understanding
- Reasoning
- Problem solving

Data Mining

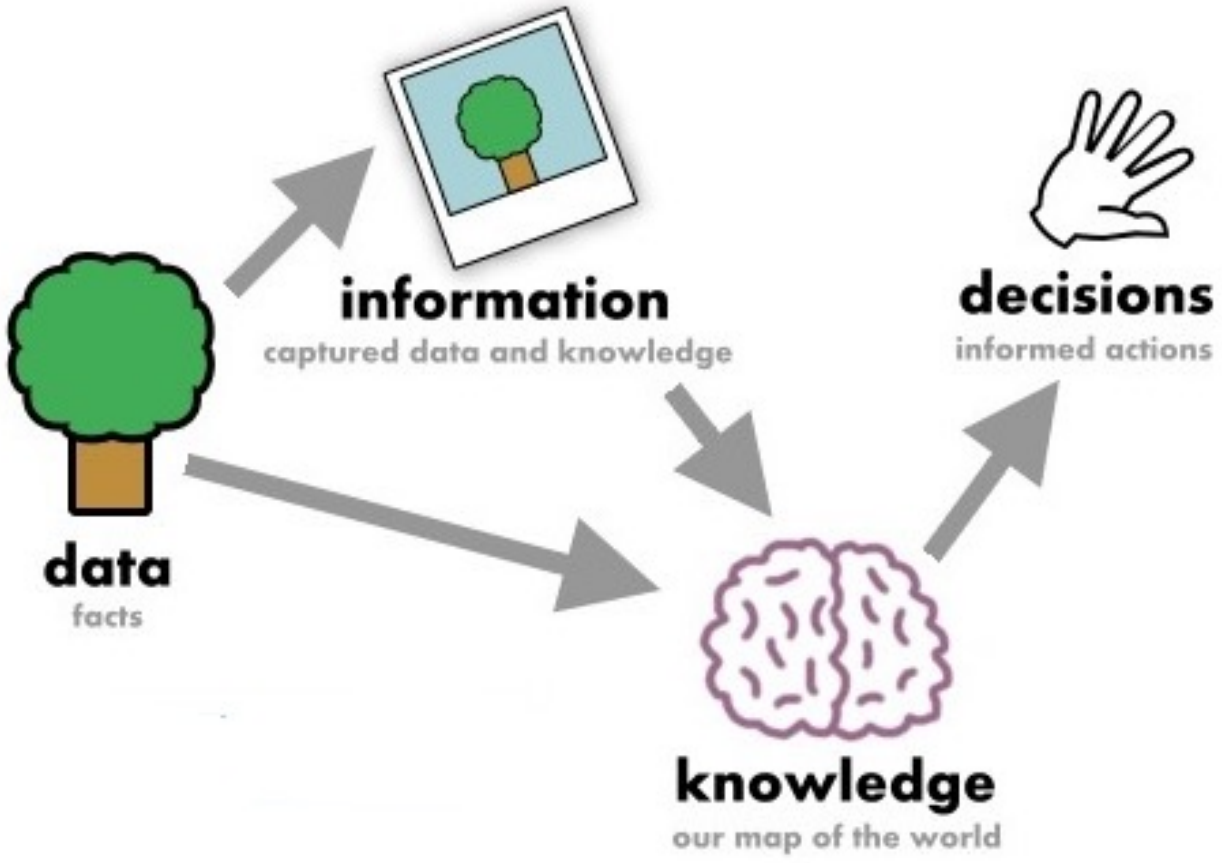
What is Data Mining

- Data mining is, in general terms, the extraction of information from data in various forms.
- What can we do with data mining?
  - Extract useful information from data sets.
  - Explore and analyse data in order to discover meaningful patterns and rules.
  - Discover meaningful correlations, patterns and trends from data.

Data Mining

- **Data**: raw un-interpreted facts
  - e.g. Tom, 20 years old, student
- **Information** relates items of Data together
  - e.g. Tom is 20 years old
- **Knowledge** relates items of Information together
  - Tom is 20 years old → Tom pays > $1500 car insurance
- **Modelling** the world (= generalising)
  - [18 - 25] years old → P(accident) = high

Data Mining


information
captured data and knowledge
decisions
informed actions
data
facts
knowledge
our map of the world

Fitting to the Business

Case 1: insurance industry

- **Problem**: AAMI insurance wanted a better and fairer way to work out premiums.
- **Approach**: The Data Mining Research Group at the Faculty of IT, Monash University used more than 30,000 customer transactions in AAMI's database to work out what characterized good and bad driver behavior.
- **Deployment**: The company could use the discovered features to set fairer premiums.

Fitting to the business

Case 2: Gambling

- **Problem:** The New South Wales TAB (gambling agency) was looking for betting patterns on horse race.
- **Approach:** Data miners looked at factors such as track conditions, weather, the number of horses running and the odds placed in each race and how they influenced betting.
- **Deployment:** More accurate prediction on the level of betting on any particular day.

DM - Machine Learning

- Supervised versus unsupervised learning.

  - Supervised Learning
    - We have training data that is annotated with known target values (labelled).
    - Our task is usually one of predicting the target values in new data: Classification and Regression.
      - Classification: predict categorical (discrete) values.
        e.g., image classification
      - Regression: predict numerical (continuous) values.
        e.g., stock price

DM - Machine Learning

- Unsupervised Learning
  - Our data is not annotated with any known target values (unlabeled).
  - Our task is usually learning the structure of the data.

    For example, we have a record of shopping habits and we want to find out whether there are any patterns of items bought together.

DM - Machine Learning

- Classifiers:
  - K-Nearest Neighbor
  - Decision Tree
  - Random Forest
  - Naïve Bayes
  - Support Vector Machine
  - Neural Network
  - ……
- Clustering algorithms:
  - K-means
  - DBSCAN
  - Agglomerative Clustering
  - Spectral Clustering
  - ……

DM - Deep Learning

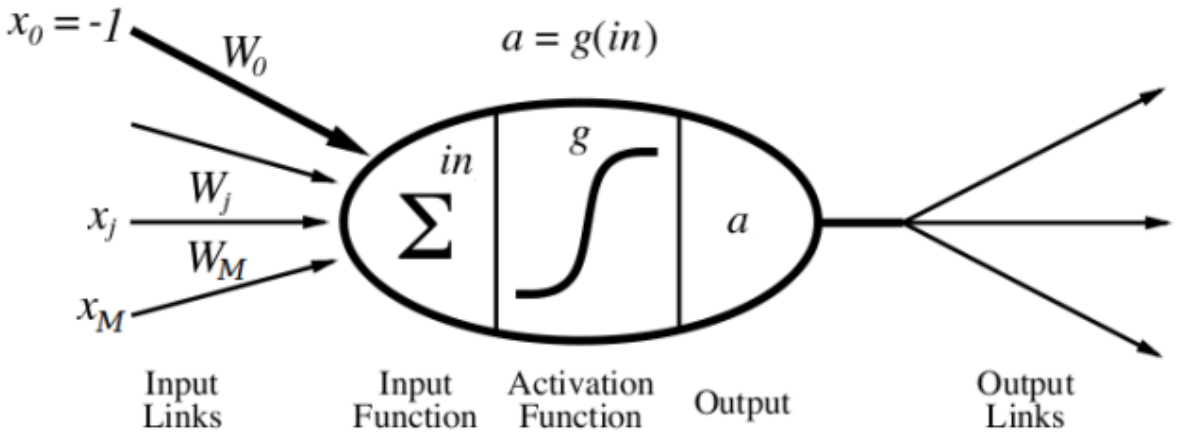
Artificial Neural Networks:

- It is inspired by human brain which is composed of cells called neurons which are interconnected to form a massive network.
- It is one of the earliest methods of AI, which attracted a lots of attention between the 1950's and 1980's due to their potential to automatically develop ways of solving problems, given appropriate training data.
- Neural networks have gone through several phases of decline and resurgence.

DM - Deep Learning

Artificial Neuron

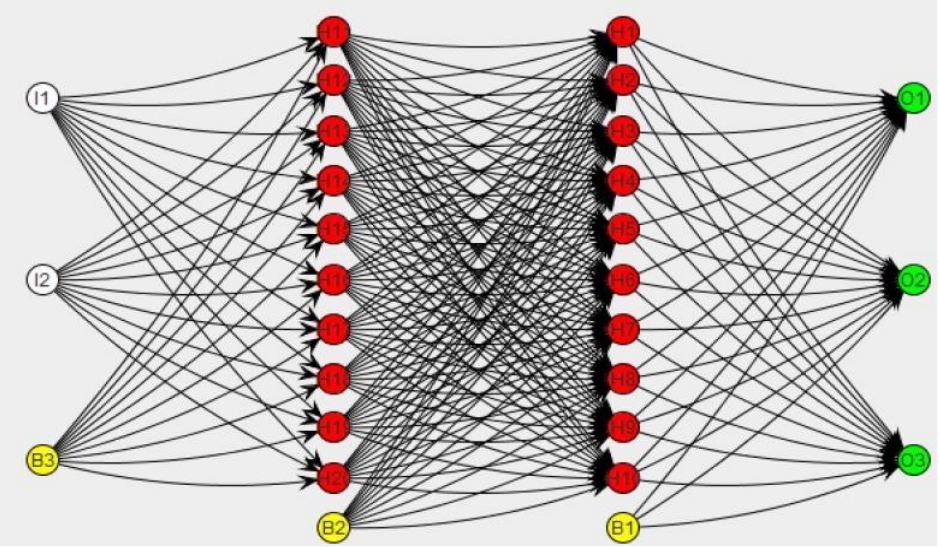
- A mathematical model of the neuron due to McCulloch and Pitts (1943):



- It is an oversimplification of real neurons, but the goal was to investigate what networks of simple units can do.
- What it does in a nutshell: It fires when a linear combination of its inputs exceeds some threshold.

DM – Deep Learning

- Neural network terms
  - Neuron
  - Connection
  - Activation
  - Layer
- Neural network training
  - Backpropagation
  - …



DM - Deep Learning

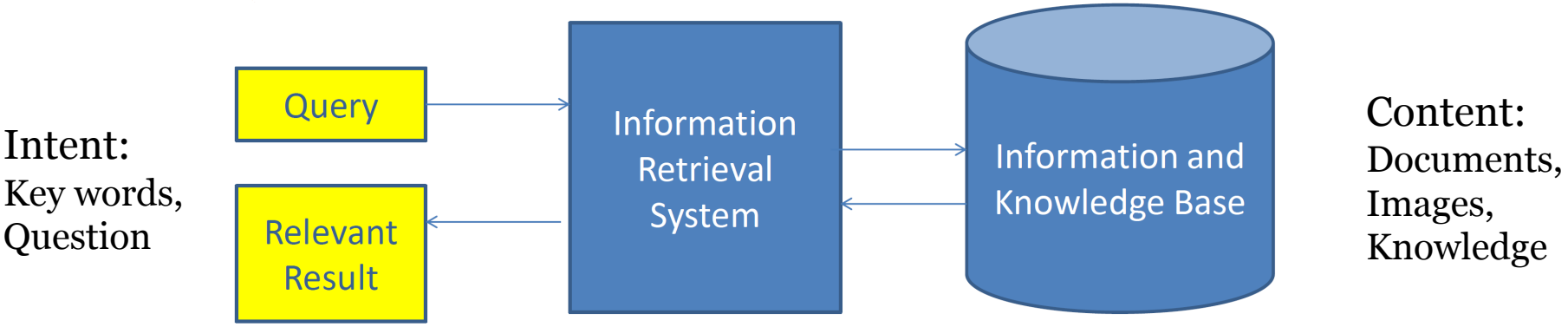
- Feed-Forward Neural Networks
- Convolutional Neural Networks
- Recurrent Networks
- Long/Short Term Memory Network
- Deep Belief Networks
- Auto-Encoders
- Generative Adversarial Network
- Variational Auto-Encoders
- …

DM – Pattern Recognition

- Pattern recognition:
  - Borrow from statistics, mathematics and signal processing.
  - Develop mathematical models largely from statistical/probabilistic frameworks for modelling data.
- Pattern recognition vs Machine learning
  - Pattern recognition is earlier
  - They have same ultimate tasks to mine patterns from data.
  - Pattern recognition fits the model of existing features, while machine learning learns features from data.

DM – Information Retrieval

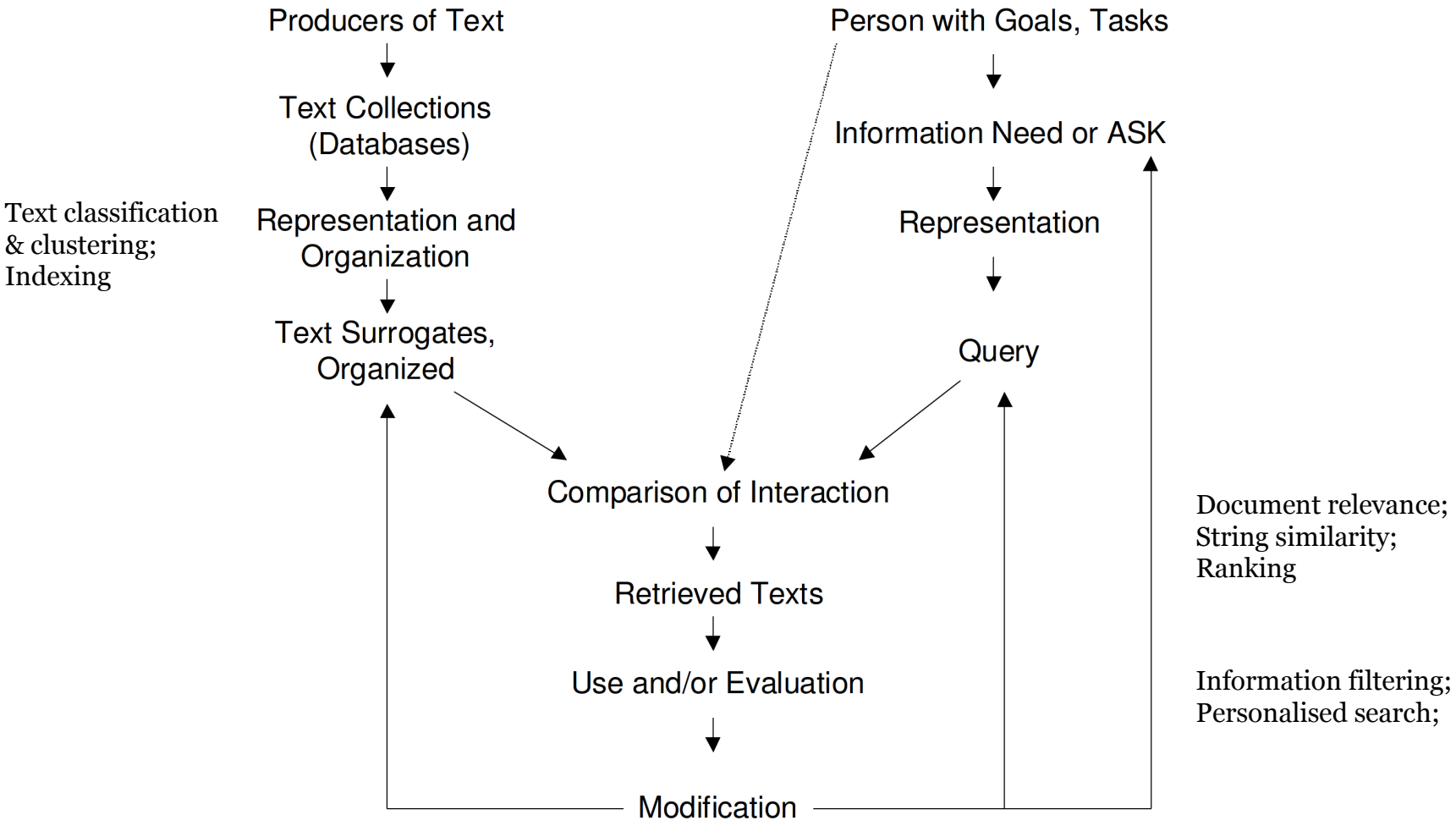
- Information Retrieval (IR) is the automatic processes that respond to a user query by examining a collection of documents and returning a sorted document list that should be relevant to the user requirements as expressed in the query.



Key Questions: How to Represent Intent and Content, How to Match Intent and Content

Source from Hang Li, SIGIR'16 Tutorial

DM – Information Retrieval



Source from Belkin 1992

DM – Information Retrieval

- Approach in traditional IR:

Document:

Star Wars: Episode VII
Three decades after the defeat of the Galactic Empire, a new threat arises.

Query:
star wars the force awakens reviews

$$q = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \xrightarrow{f(q,d)} d = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

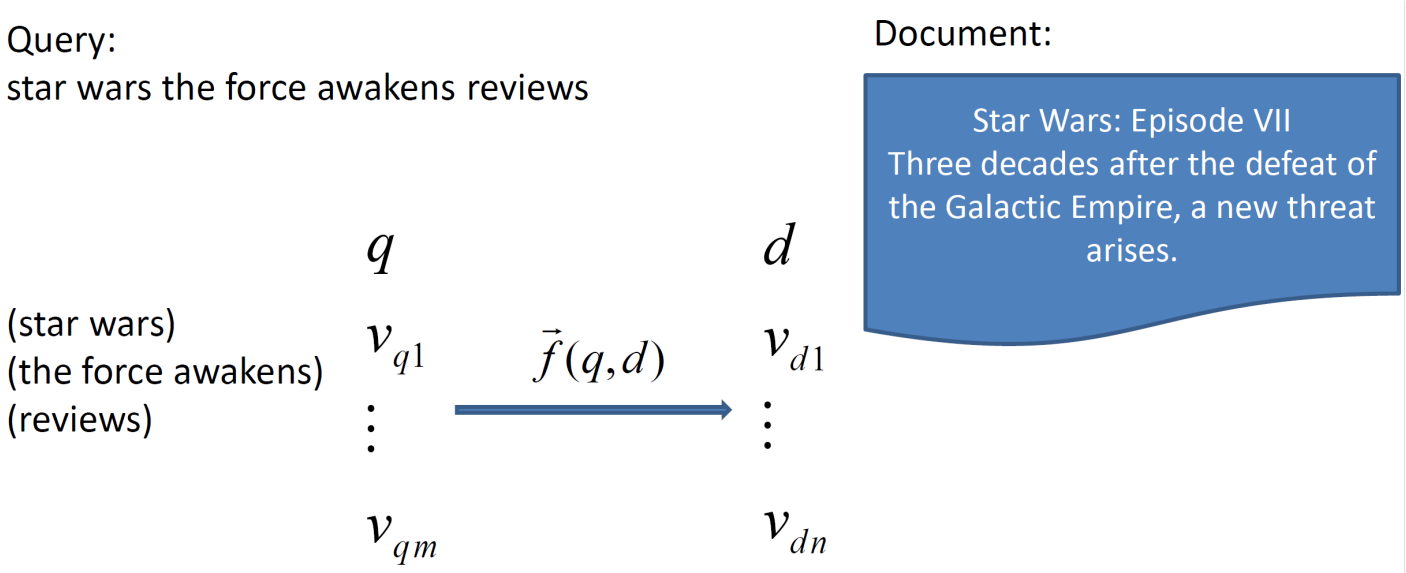
$$f_{VSM}(q,d) = \frac{\langle q,d \rangle}{\| q \| \cdot \| d \|}$$

- Representing query and document as word vectors
- calculating cosine similarity between them

Source from Hang Li, SIGIR'16 Tutorial

DM – Information Retrieval

- Approach in Modern IR:

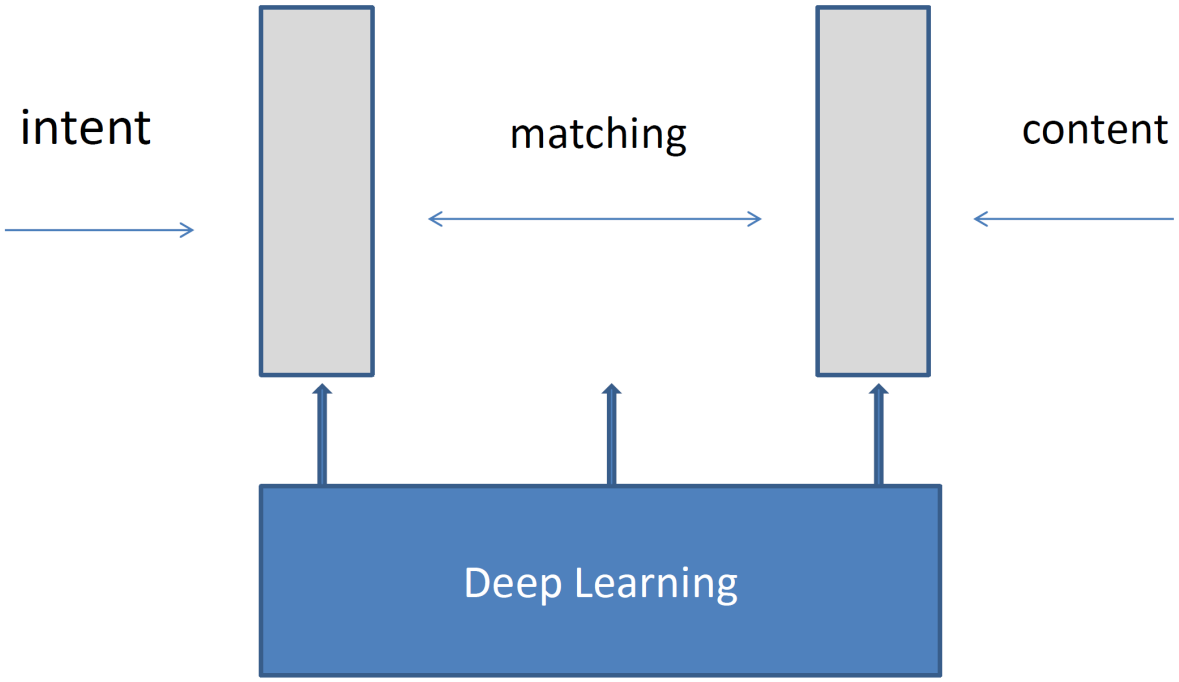


- Conducting query and document understanding
- Representing query and document as multiple feature vectors
- Calculating multiple matching scores between query and document
- Training ranker with matching scores as features using learning to rank

Source from Hang Li, SIGIR'16 Tutorial

DM – Information Retrieval

- Representation and Matching are key problems in IR



Recent Progress: Deep Learning Enables Representation Learning and Matching in IR

Source from Hang Li, SIGIR'16 Tutorial

Language Understanding

Language Understanding

- Understanding depends on
  - the result of parsing,
  - the lexical information,
  - the actual context, and
  - commonsense reasoning.

Language Understanding

- Part-of-Speech Tagging
- Syntactic Parsing
- Dependency Parsing
- Semantic Role Labelling

Language Understanding

- Part-of-Speech Tagging

  is the process of assigning a part-of-speech marker to each tagging word in an input text.

  Example:

  Input: My aunt's can opener can open a drum

  Output: My/PRP$ aunt/NN 's/POS **can/NN** opener/NN **can/MD** open/VB a/DT drum/NN

  Tags' meaning in this example:

  - PRP$: possess. Pronoun
  - POS: possessive ending
  - NN: sing or mass noun
  - MD: modal
  - VB: verb base form
  - DT: determiner

Language Understanding

- Part-of-Speech Tagging
  - Rule-based methods by applying manually constructed rules (Kučera and W. Nelson Francis, 1960s)
  - Probability methods:
    - Hidden Markov Model (HMM)
    - Maximum Entropy Markov Models
    - Conditional Random Fields
  - Neural Network methods:
    - Bi-LSTM (Plan et al. 2016)
    - …

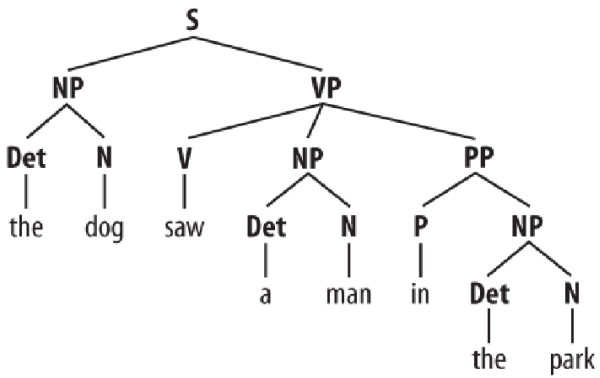
Language Understanding

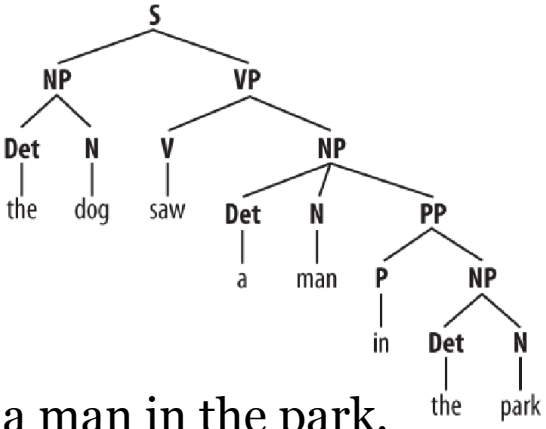
- Syntactic Parsing
  - Syntactic parsing is the task of recognizing a sentence and assigning a syntactic structure to it.
  - Probabilistic models:
    - Cocke-Kasami-Younger (CKY) algorithm
    - Probabilistic lexicalized CFGs

Context free grammar:

S -> NP VP
NP -> Det N | Det N PP
VP -> V NP | V NP PP
PP -> P NP
Det -> "the" | "a"
N -> "man" | "dog" | "park"
V -> "saw"
P -> "in"

Syntax Trees:



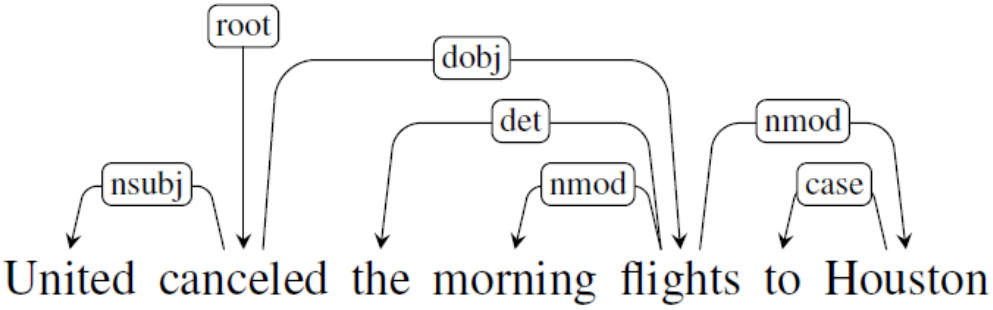


The dog saw a man in the park.

Language Understanding

- Dependency Parsing:

  is the task of extracting a sentence's grammatical structure and defines the relationships between "head" words and words, which modify those heads.



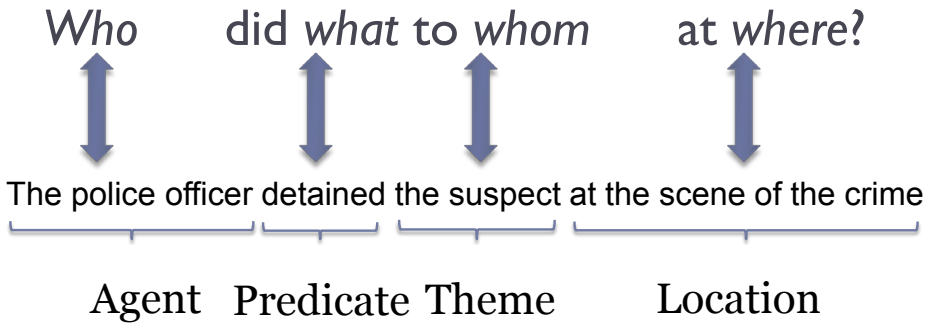
The clausal relations NSUBJ and DOBJ identify the subject and direct object of the predicate cancel, while the NMOD, DET, and CASE relations denote modifiers of the nouns flights and Houston.

Language Understanding

- Dependency Parsing
  - Transition-based dependency parsing
  - Graph-based parsing
  - Neural Dependency Parser
- Universal Dependencies treebanks (link)
  - is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages.
  - UD is an open community effort with over 300 contributors producing more than 150 treebanks in 90 languages.

Language Understanding

- Semantic Role Labelling
  - is the process that assigns labels to words or phrases in a sentence that indicate their semantic role in the sentence, such as that of an agent, goal, or result.
  - Semantic role express the abstract role that arguments of a predicate can take in the event.



Source from Jurafsky & Martin, 2019

Language Understanding

- Semantic Role Labelling
  - Rule-based (1970s-1980s)
  - Learning-based
    - Pruning-> identification->classification
  - Chucking task:

    Formulate Chunking as I (inside) – O (outside) –B (beginning) tagging (classification), e.g.,:

The	*morning*	*flight*	*from*	*Denver*	*has*	*arrived*
B_NP	I_NP	I_NP	B_PP	B_NP	B_VP	I_VP

Reasoning

Reasoning

- Reasoning aimed at generating answers to unseen questions by manipulating existing knowledge with inference techniques.
  - Example: John is either in the car or in the house. He isn't in the car, therefore he is in the house.
- Two components:
  - Knowledge, such as a knowledge graph, common sense, rules, assertions extracted from raw texts, etc.;
  - An inference engine, to generate answers to questions by manipulating existing knowledge.

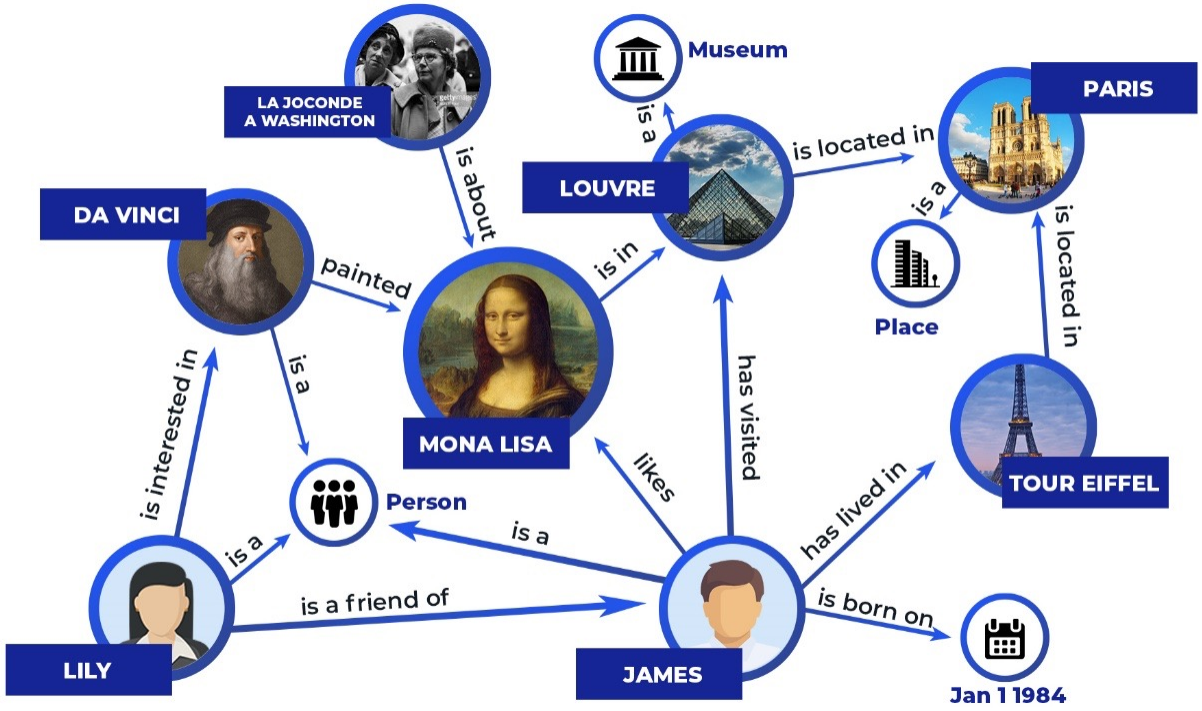
Reasoning

- Knowledge Graph
  - A knowledge graph acquires and integrates information into an ontology and applies reasoning to derive new knowledge.
  - Google knowledge graph.
    Freebase

Reasoning

- Knowledge Graph
  - Edges, Vertices and Relations

  (Da Vinci, painted, Mona Lisa)



source

Reasoning

- Knowledge Graph
  - Research
    - Representation: logic, n-tuples, database
    - Knowledge acquisition and Construction: Named Entity Recognition, relation extraction.
    - Temporal knowledge graph

Reasoning

- Inference engine
  - Integer linear programming (ILP)
  - Probabilistic method: e.g., Bayesian Networks, Markov logic networks (MLNs)
  - Neural methods: Memory network and variants.

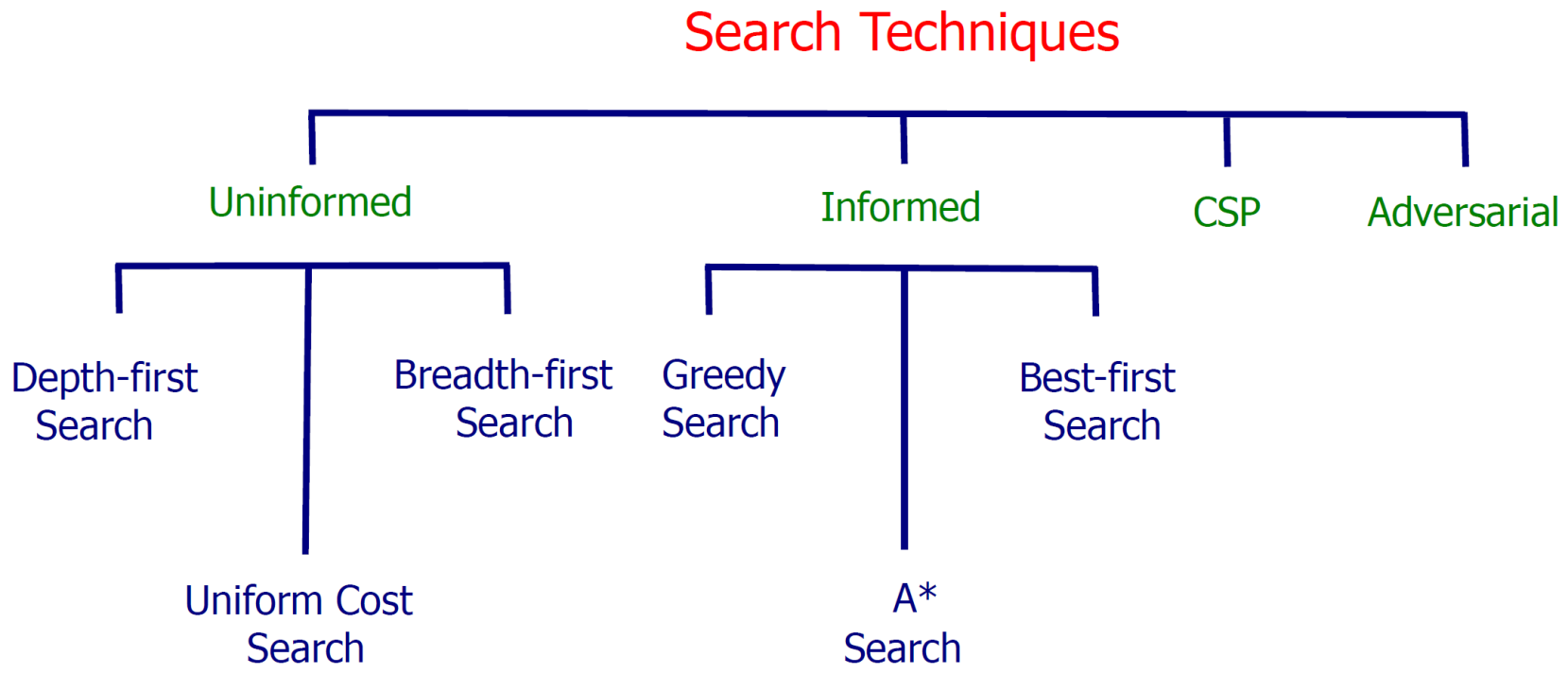
Problem Solving

Problem Solving

- Problem solving by Search
  - searching in an internal representation for a path to a goal.
  - Differs with search the world or search on the Web.

Problem Solving

- Problem solving by Search



Problem Solving

- Problem solving by Search
  - Genetic Algorithm
    - Inspired by the process of natural selection
    - Belongs to evolutionary algorithms
    - Used to generate solution in search problems
    - Mutation, crossover, selection

Fields that influence(d) AI

Philosophy	Logic, methods of reasoning, mind as physical system foundations of learning, language, rationality
Mathematics	Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability
Economics	Utility, decision theory
Neuroscience	Physical substrate for mental activity
Psychology	Phenomena of perception and motor control, experimental techniques
Computer engneering	Building fast computers
Control theory	Design systems that maximize an objective function over time
Linguistics	Knowledge representation, grammar

References

- COM329, MQ by Dr. Jia Wu
- SIGIR tutorial 2016 by Prof. Hangli