

Progetto di Ingegneria Informatica

Proposte di progetto

Prof. Letizia Tanca

These slides contain:

- General research areas for projects
- After a slide on the research area, one or more examples of current specific subjects within one of these areas
- More specific subjects come up while the research goes on, and will be presented to you when you come to ask for a project

Some simple pieces of advice

- You can ask for a project when you have time to work on it
- Depending on the character of the student(s) and on the phase of the research, a project can range from a very technical one, implementation-oriented, to a more conceptual one, aimed at finding a principled solution (model, algorithm...) to a problem

Heterogeneous Data Integration

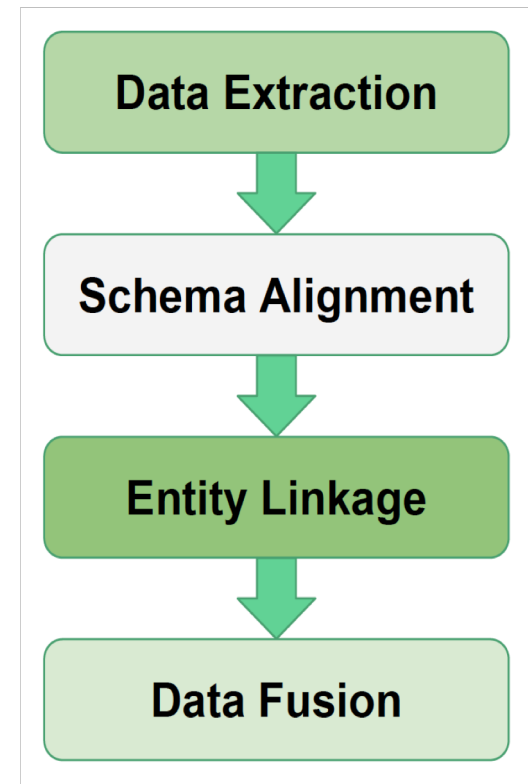
Theses may be offered on any of the Data Integration phases (depending on the period in which the students is willing to work)

Data extraction: extracting structured data; more difficult when non-relational data exist

Schema alignment: aligning types and attributes; helpful when different relational schemas exist

Entity linkage: recognizing and linking pairs of records or entities referring to the same objects in different data

Data fusion: resolving conflicts; necessary in presence of different values for the same piece of data



Data Integration and Machine Learning: A Natural Synergy

Xin Luna Dong @ Amazon.com
Theo Rekatsinas @ UW-Madison
<http://dataintegration.ai>

<http://www.vldb.org/pvldb/vol11/p2094-dong.pdf>

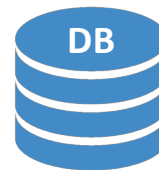
https://thodrek.github.io/di-ml/vldb2018/slides/diml_vldb2018.pdf

<https://www.youtube.com/watch?v=8RMSaFVPT1Q&t=3s>

Query Reverse Engineering

Expert user
of DB

Query
select *
from Student
where Name="John"



ID	Name	Age
1	John	19
2	John	37

Non-expert
user of DB

ID	Name	Age
1	John	19
2	John	37



**Query Reverse
Engineering
Algorithm**

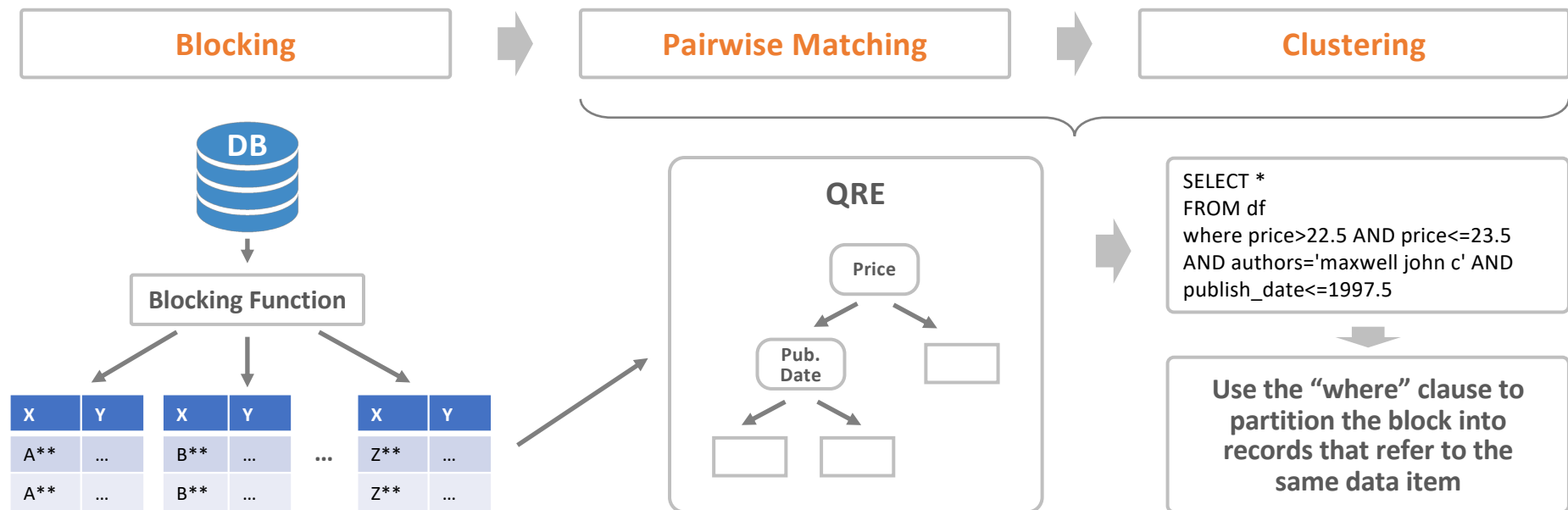
Query n

Query 2

Query 1

select *
from Student
where Name="John"

Entity resolution with QRE



Preliminary results: Precision 95% (Book Dataset with 10k rows)

Next steps: improve robustness, test on other datasets

Personalization and Context-Awareness

Data Pervasivity and Personalization

- Rapid development of the web: the web has enabled people with varied goals and characteristics to access an ever-growing amount of information and online services.
- Dominance of hand-held electronic devices such as tablets and smartphones: these have made information access possible from anywhere and at any time.
- The massive data generated by devices, applications, and people creates overload but also enables new, improved applications and services.

→ PERSONALIZATION as a meaningful way to sort out choices and present them in useful ways to users taking into account their characteristics, such as their current context

→ The CONTEXT is the who, what, when, where, and why in which an action takes place. A context-aware framework (CAF) attempts to represent the awareness of these 5 Ws in an intelligent and reactive computing system

Context-Awareness in accident prevention

The RECKON Project: sensing context for safety protection and accident prevention



Project Goals



**Context-aware data gathering,
integration and analysis**

- Increase the accuracy of detection and diagnosis
- Improve preventive actions
- Better understand the needs of solutions
- Cut follow-up costs



Credits: Beng Chin OOI –VLDB 2018

Ethics in Data Management

Claim: Big Data is algorithmic, therefore it cannot be biased! And yet...

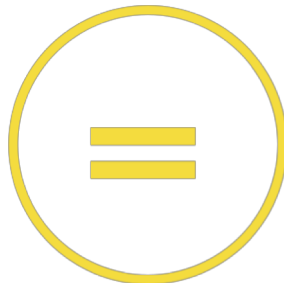
- Algorithms discriminate just like humans do, but at a larger scale
- Processes are opaque, and defy public scrutiny
- It is our responsibility to understand the issues and offer technological solutions that address them
- Datasets can embed prejudice: Garbage in garbage out!

Technology must be informed by ethical and legal considerations.

Data Protection



Neutrality



Transparency



Diversity



Fairness

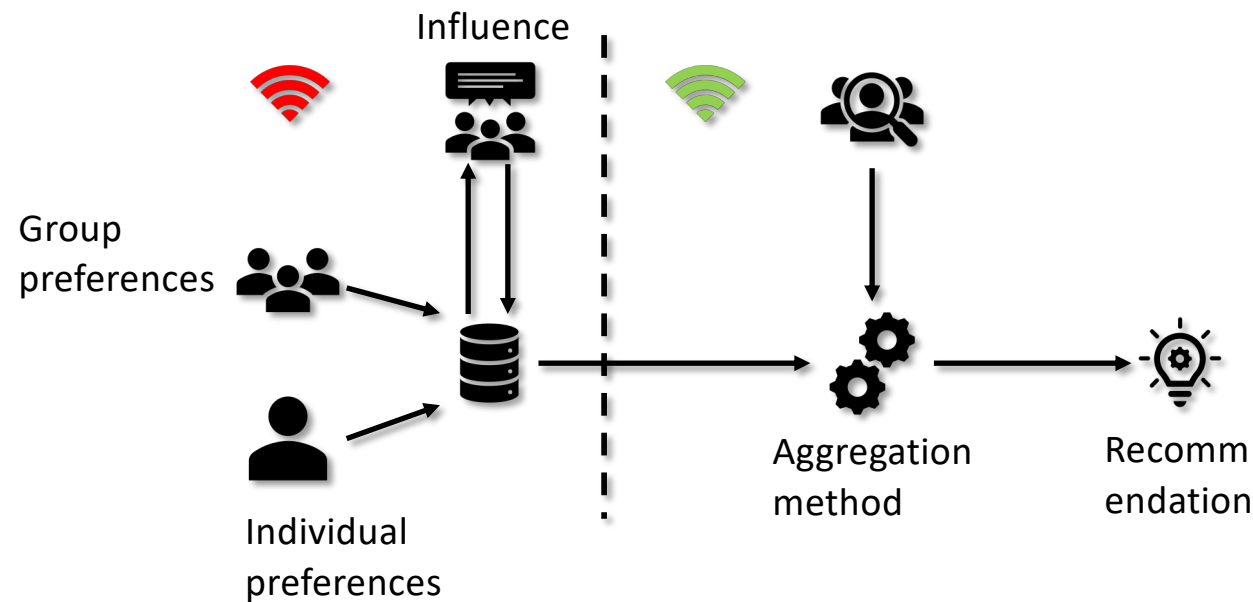


A Context-Aware Group Recommender System

Key Idea: Use Influence to aggregate individual scores

$$Score(G, i, c) = \frac{\sum_{u \in G} infl(u, c) \cdot score(u, i, c)}{\sum_{u \in G} infl(u, c)}$$

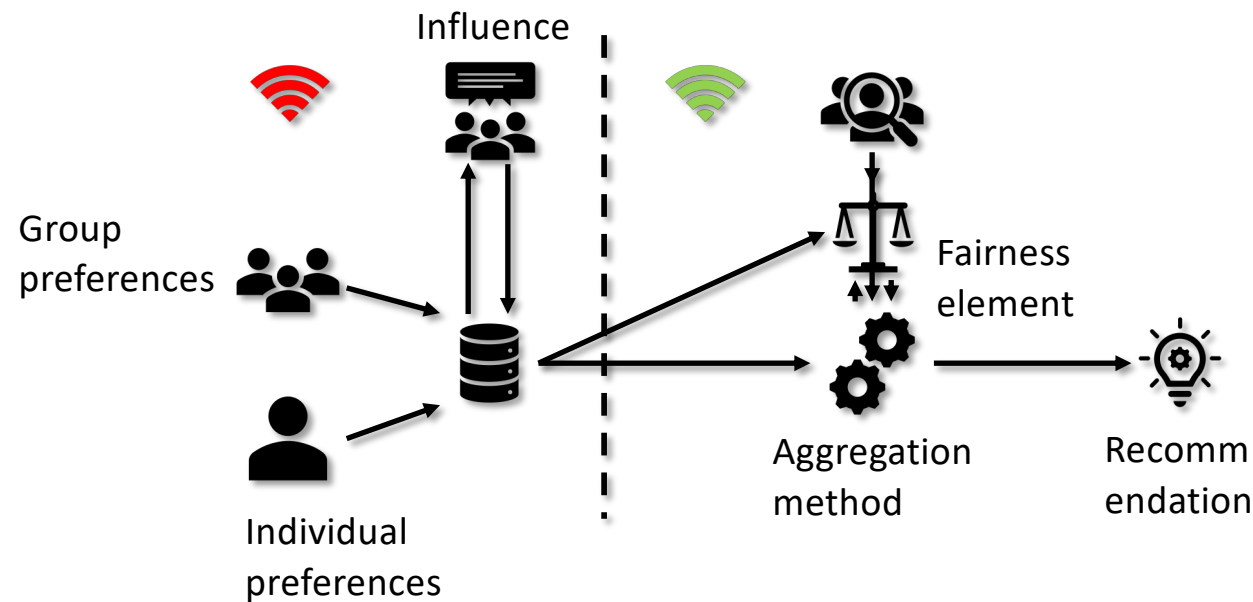
G : group
 i : item
 c : context
 u : user
 $infl$: influence



A Fair Context-Aware Group Recommender System

Key Idea: Influence is the opposite of Fairness!

$$Score(G, i, c) = \frac{\sum_{u \in G} infl(u, c) \cdot score(u, i, c)}{\sum_{u \in G} infl(u, c)} \cdot Consensus(G, i, c)^{size}$$



G: group
i: item
c: context
u: user
infl: influence

Urban diagnostics





Urban diagnostics

*In IMM, **urban diagnostics** is a model-based approach to define, through a rigorous qualitative and quantitative representation, the state of a system and its performance. By investigating urban context as a Complex Adaptive System, IMM analyzes patterns of problems and malfunctioning conditions to infer the source of the problem.*

The IMM project contributes to the implementation of the following Sustainable Development Goals:



Electricity Demand

Electricity Demand Forecasting for Buildings

The goal of this thesis is to develop a new algorithm for forecasting building's power demand and comparing the proposed method to the best scoring approaches present in the literature on the provided datasets.

Datasets:

- 2 years of power consumption of a building. A different time series is available for each of the tree floors of the building under consideration. For the same period, temperature and humidity readings (of an ARPA weather station placed near the building) are available.
- 1 month of power consumption of 1449 buildings. For the same period, weather data are available.
- Other datasets proposed by the student are welcome!

Electricity Demand Forecasting of Buildings

Requirements:

- This thesis can be jointly done by two students.
- This is a coding-intense thesis! Hence, strong programming skills (python and/or r) are required.
- Knowledge (at least basic) of Deep Learning models for sequence data (e.g., RNNs, LSTMs) and how to implement them in Keras/TensorFlow is also needed.

Recommended readings:

- Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence-based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews.
- Almalaq, A., & Edwards, G. (2017). A review of deep learning methods applied on load forecasting. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA).
- Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews.