

Theoretical Hackaton Whiteboards

Group 5

Francesco, Mariia, Lyudmila, Joel, Een, Peter, Sanda,
Francesco, Mihai

NAME

ACCESS TO
DATA SOURCES

QUERY

EVALUATION

MARIA

CRAWLER

INDEXING algorithm

SPARQL query

RDF annotations

IR: TF-IDF...

RELEVANCE JUDGEMENTS

LYDMILA

ONTOLOGIES

AUTOMATIC RANKING
HUMANS

JOEL

UNSTRUCTURED
(TEXT PROCESSING)

SVD

FREE TEXT
QUERY

GOLD STANDARD

EEN

SENSOR DATA (HEALTHCARE)
OPEN EHR, LHM
PERSONAL INFORMATION
HETEROGENEOUS

ARCHETYPE QUERY
PATTERNS

ASK THE DOCTOR.

PATIENT RECORDS/HISTORY

BRAIN COMP. INTERFACE

PETER

DATA TABLE w ATTRIBUTES

UNSTRUCTURED → STRUCTURED

FORMAL CONCEPT ANALYSIS
SELECT
FUZZY SET

QUALITY MEASURES (STABILITY OF CONCEPTS)

NUMERIC

- USER STUDIES → GOLD STANDARD

SAUDA

MULTIMODAL CORPORA / LOW DIM.
PREPROCESSING REPRESENTATION
ANNOTATIONS

VECTOR DISTANCE

GOLD STD.

USER EXPERIENCE - POWER

EYE TRACKING

FRANCESCO

~~DB~~ + scalability
(RDF, Relational)

GRAPH-BASED
SCHEMA-BASED TRANSFORMATION

FREE TEXT → SQL

(DISAMBIGUATION!!)
MATCHING TABLES

TIME, form a efficiency
3! Benchmark: 50 queries
Information need

MIHAI

UNSTRUCTURED DOCU CO.
ANNOTATIONS

INFORMATION NEEDS
KEYWORDS

BENCHMARKS, TEST COLLECTIONS

Challenges

NAME

MARIA

PERSONALIZATION
(HUMAN OR NOT)

LYDMILA

RELEVANCE
JUDGEMENTS

JOEL

INFORMATION NEED
& QUERY

EEN

CONTEXT
(REF. & USE)
TIME-SERIES

PETER

REDUCING TIME
OUTPUT
(RANKINGS)

SANDA

MULTILINGUALITY
CONTEXT

FRANCESCO

INFORMATION
FUSION
- WHAT DOES THE
USER WANT?

MIHAI

SEMANTICS

ACCESS TO
SOURCES

QUERY

EVALUATION

SPARQL query

RDF annotations

IR: TF-IDF...

RELEVANCE JUDGEMENTS

AUTOMATIC RANKING
HUMANS

GOLD STANDARD

SVD FREE TEXT
QUERY

ARCHETYPE QUERY
PATTERNS

ASK THE DOCTOR.

PATIENT RECORDS/HISTORY

BRAIN COMP. INTERFACE

FORMAL CONCEPT ANALYSIS

SELECT

FUZZY SET

QUALITY MEASURES (STABILITY OF CONCEPTS)

NUMERIC

- USER STUDIES → GOLD STANDARD

VECTOR DISTANCE

GOLD STD.

USER EXPERIENCE - POWER

EYE TRACKING

GRAPH-BASED

SCHEMA-BASED TRANSFORMATION

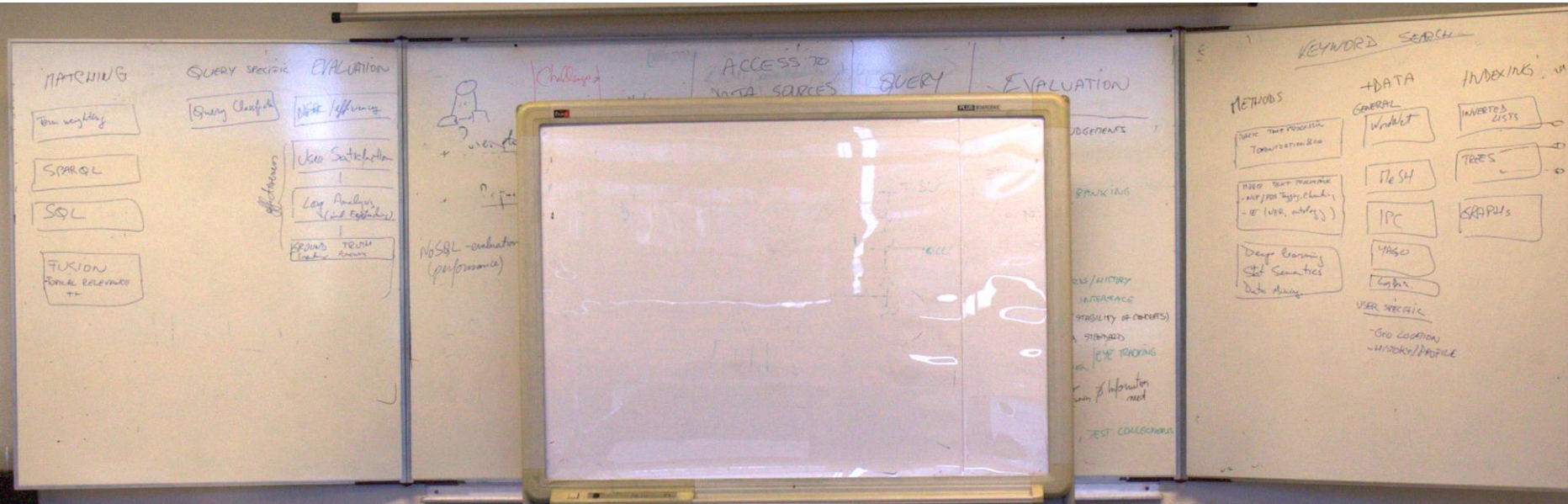
FREE TEXT → SQL

(RECONFIGURATION!!)
MATCHING TABLES

TIME, form a efficiency
3! Benchmark: 50 queries
Information need

BENCHMARKS, TEST COLLECTIONS

INFORMATION NEEDS
KEYWORDS



KEYWORD SEARCH

METHODS

Basic Text Processing
Tokenization & Co

More Text Processing
- NLP / POS Tagging, chunking,
- IE (NER, ontology)

Deep Learning
Stat Semantics
Data Mining

+ DATA

GENERAL

WordNet

MeSH

IPC

YAGO

Lybra

USER SPECIFIC

- Geo Location
- History / Profile

INDEXING

INVERTED LISTS

TREES

GRAPHS

MATCHING

Term weighting

SPARQL

SQL

FUSION

Topic Relevance

++

QUERY SPECIFIC EVALUATION

Query Classification

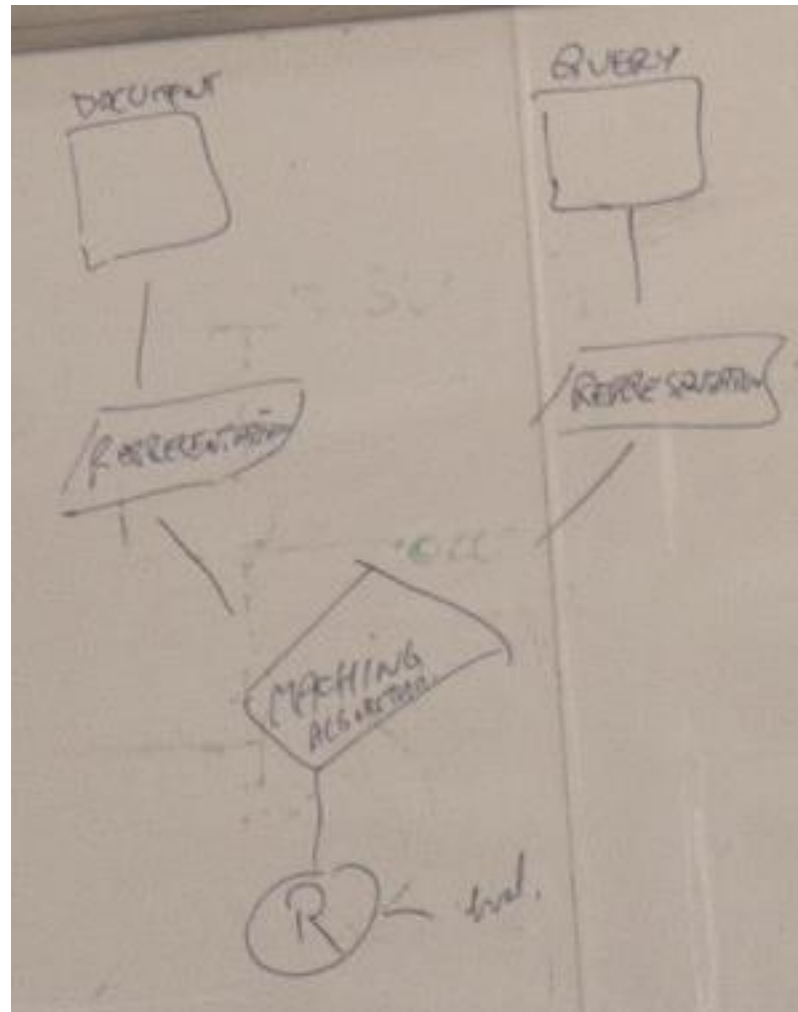
Cost / Efficiency

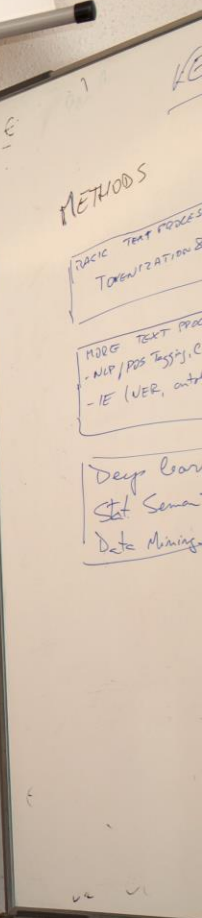
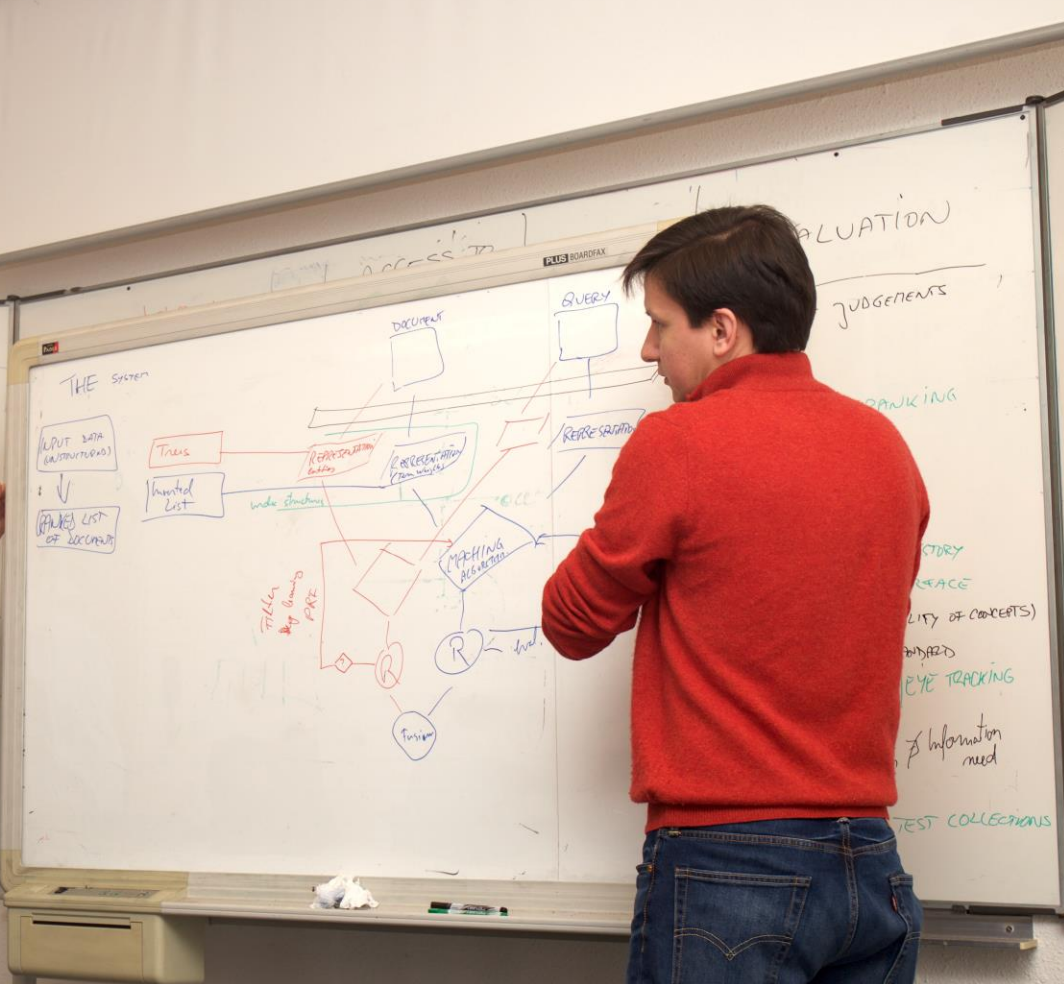
effectiveness

User Satisfaction

Log Analysis
(incl. Expectations)

GROUND TRUTH
(reality Answer)





I. Query

INPUT: TEXT
USER MODEL (Text)

OUTPUT: RANKED DOCUMENTS

I.1. text: basic processing $\rightarrow$ tokens
: entity recognition $\rightarrow$ entities

I.2. matching: 1)

score_{ed} = ...

I.3 if (score_e > 0): PRF $\leftarrow$ PRF $\cup$ {0}

I.4 $t_{exp} \leftarrow$ Term Weighting (PRF)

$$\text{score}_{t_0} = \lambda \cdot w_{t_2} + \beta w_{t_{um}} + \gamma w_{t_{exp}} + (1 - \lambda - \beta - \gamma) w_{t_c}$$

I.5 $\text{score}_y = \sum_{t \in y} \text{score}_{t_0}$

I.6. RANK

II. INDEX CREATION

INPUT: DOCUMENT

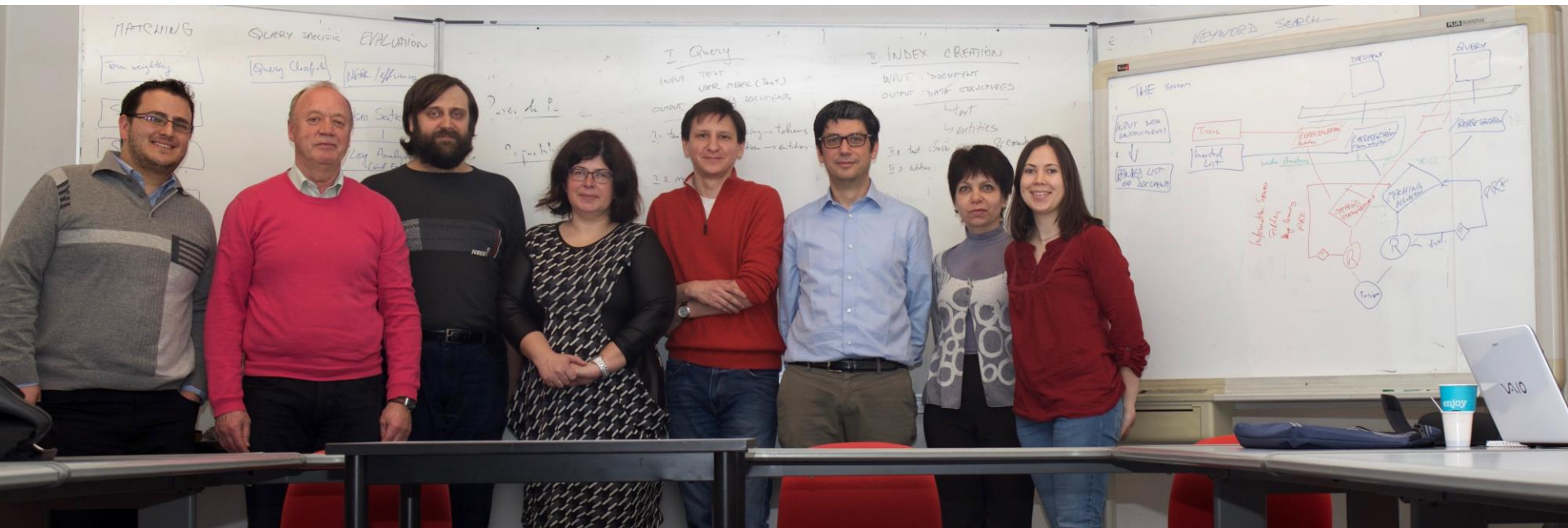
OUTPUT: DATA STRUCTURES

$\hookrightarrow$ text

$\hookrightarrow$ entities

II.1. text: Basic processing & counting $\rightarrow$ 1.

II.2. entities: 1. 1.



Thank you