



TIP@50 analysis methodology based on HSE intelligentFOResightAnalytics (iFORA) system

ikuzminov@hse.ru



dmeissner@hse.ru



pbakhtin@hse.ru



lgokhberg@hse.ru



OECD, CSTP – TIP Workshop

Paris, 12-13 March 2018



Outline

- Sources of data and main text mining approaches
- TIP@50 analysis methodology
- Software decisions, strengths and challenges
- Application of methodology to other policy studies
- TM requirements for usage in other policy studies



Sources of data and main text mining approaches

OECD TIP documents

agendas

≈90 documents

summaries

≈90 documents

reports

≈100 documents



iFORA

Computer linguistics

Big data, AI & machine learning

Ontologies and expert knowledge

unique combination of three text mining approaches



TIP@50 analysis methodology

Text segmentation

Sentence tokenization

Lemmatization, stemming, morphological analysis

Syntactic analysis

Word and document embeddings, topic modeling

Entity linking

Statistical analysis, network analysis, clustering



sentence 1
sentence 2
sentence 3

sentence

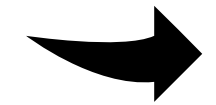
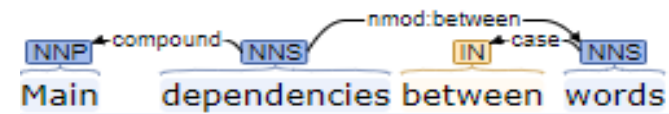


word 1
word 2
word 3

word



lemma
stem
part of speech
...



terms

term /
document

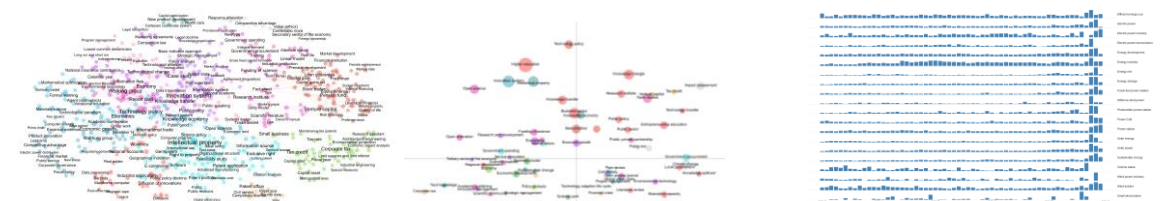


$vector(x_1, x_2, \dots, x_n)$

term (vector)



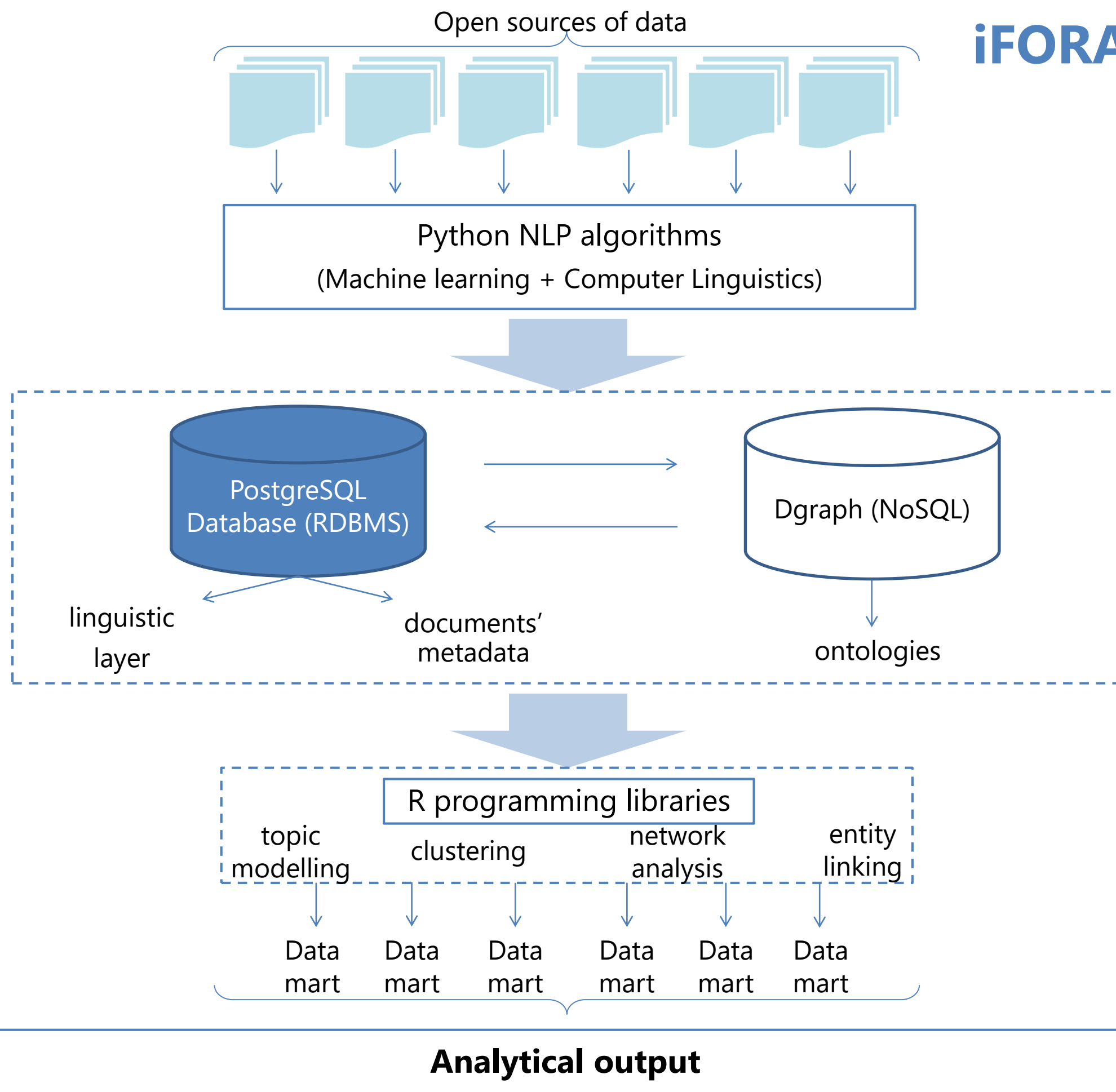
entity (vector)





Software decisions, strengths and challenges

iFORA



Strengths

- + open source libraries
- + open source data
- + flexible in-house development
- + high-performance local servers
- + state-of-art algorithms

Challenges

- ! hardware costs
- ! high quality human resources costs



Application of methodology to other policy studies

- Foresight of future markets, emerging technologies and competences
- Identification of centers of excellence and key players
- Evidence-based policies: elaboration and evaluation
- Projects expertise and selection



TM requirements for usage in other policy studies

- Standard protocols and procedures for reproducibility and comparability reasons
- General applicability to any policy study
- Identification of causalities and interrelations
- Linkages to open science databases and protocols



Thank you for your attention!

<http://issek.hse.ru>

<http://foresight.hse.ru>

<https://prognoz2030.hse.ru/>

ikuzminov@hse.ru



pbakhtin@hse.ru



dmeissner@hse.ru



Igokhberg@hse.ru

