Linguistic Semantics for Search Precision and Recall Improvement (part 2)

Ilya Tikhomirov

RUSSIAN ACADEMY OF SCIENCES. INSTITUTE FOR SYSTEMS ANALYSIS.
117312, Moscow, prospekt 60-let Oktyabria, 9, ISA RAN phone/fax +7 (499) 1350463, E-mail: matandra@isa.ru www.isa.ru

Introduction

The second lecture gives an introduction to Heterogeneous Semantic Networks. We take into consideration text representation model in semantic search tasks: how Communicative Grammar and Heterogeneous Semantic Networks can be used for search precision and recall improvement.

Definition

Semantic network is a network which represents semantic relations between the concepts.

Semantic networks can be used for knowledge presentation or description of reality.

Text semantics can be formalized by semantic networks.

History

"Semantic Nets" were first invented for computers by Richard H. Richens in 1956 as an "interlingua" for machine translation of natural languages.

Heterogeneous semantic networks (1)

Heterogeneous Semantic Network is the family of oriented graphs, which have same sets of vertices, and each edge have some interpretation procedure.

Heterogeneous Semantic Networks were invented by Gennady S. Osipov in 1986.

Heterogeneous semantic networks (2)

Heterogeneous semantic networks is an algebraic system:

- $\boldsymbol{W}$=(**D**, **S**, **G**, **R**, **F**),
- **S** – set of names of objects,
- **R** – a family of relations on $\boldsymbol{S}x\boldsymbol{S}$,
- **D** – a universe of sets $\{D_1, D_m, \ldots, D_n\}$, where each D_i is said to be the set of attributes,
- **G** – set of subsets *g* of tuples in Cartesian product $D^k=D_{i1}xD_{i2}x\ldots xD_{ik}$, which are related for each name *s* from **S**, also called *extensional* of *s*.
- **F** – a family of functions $\{f_1, f_2, \ldots, f_m\}$, which maps D^k into some of sets D_i from **D**.

Heterogeneous semantic networks (3)

- The sentence on natural language maps into heterogeneous semantic network with objects (entities) as vertices and semantic relations between objects as edges.

Semantic relations

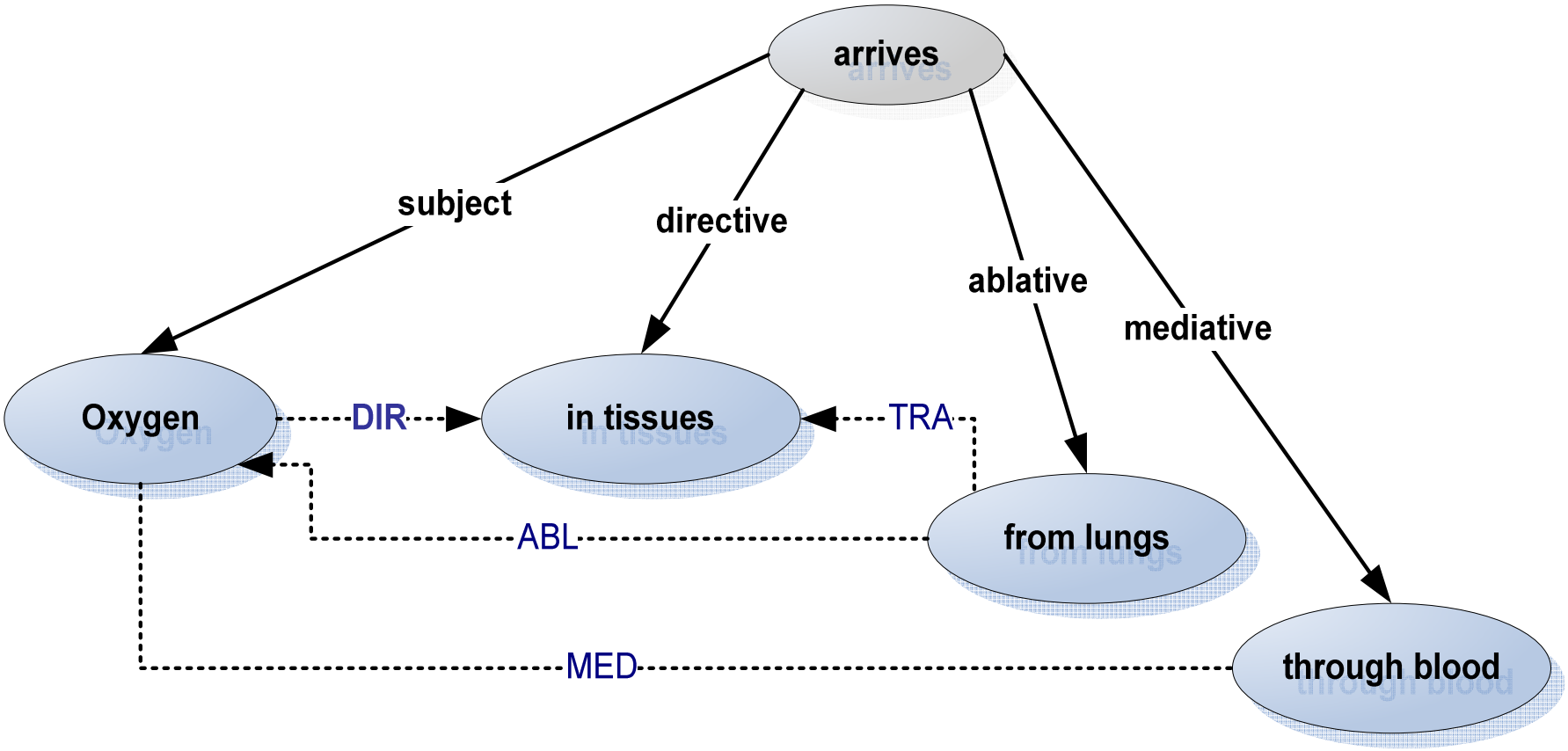
Semantic relation is an relation on the set of syntaxemes values.

Relation examples:

- Des - one component denotes destination of an other component;
- Dir - one component denotes direction of an other component;
- … (near 65).

Heterogeneous semantic networks (4): Example

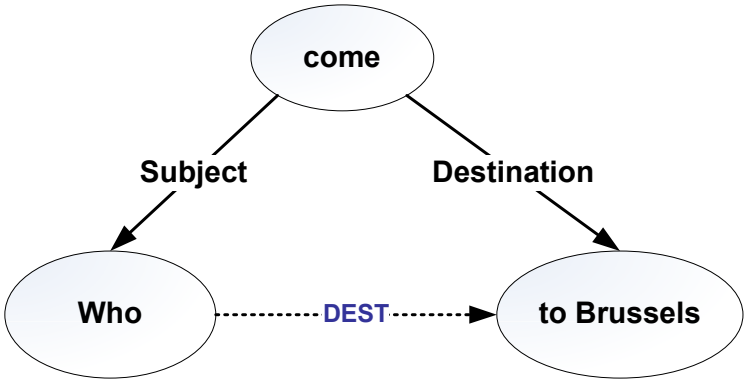
Oxygen arrives in tissues from lungs through blood



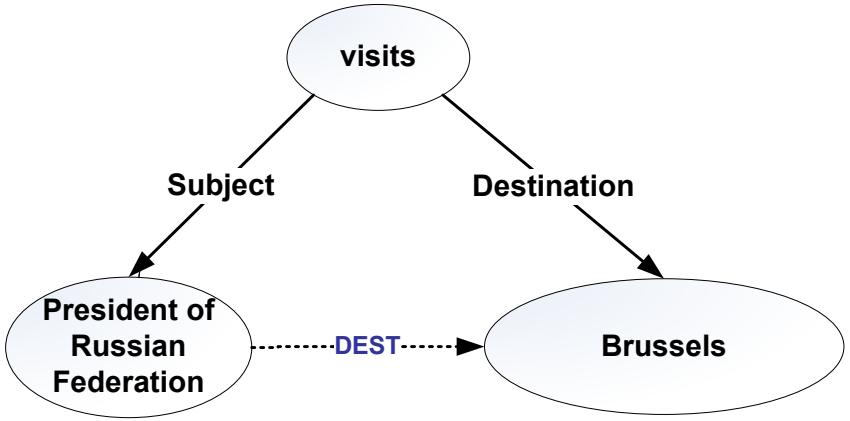
Semantic search (1)

- The main idea of semantic search is semantic matching of user query with searched documents.
- Semantic search involves generation of semantic structures (images) of documents and queries.
- The semantic structures of documents are stored in semantic indices.

Heterogeneous semantic networks: Example



Semantic network of query



Semantic network of document

Semantic search (2)

- The semantic image is presented as the semantic network so the semantic matching of query and document consists in comparison of corresponding networks vertex by vertex and relation by relation.

Semantic search (3)

Semantic search extends standard statistical approaches with the extra information resulting from the linguistic processing of texts.

The rank of the document is calculated according to 4 levels of information:

- Terms (based on TF*IDF, **well-known!).**
- Syntactic dependencies (**well-known!**).
- Semantic meanings.
- Semantic relations.

Semantic search (5)

How to get Semantic meanings and Semantic relations from the text?

Text analysis:

- Lemmatizing.
- Morphological analysis.
- Syntactic analysis.
- **Semantic analysis:**
  - **Semantic dictionary.**
  - **Linguistic rules.**

Semantic dictionary

Semantic dictionary contains predicate words and description of the semantic meanings and relations.

Example: Mary loves John for the kindness.

Verb= **love**

Meaning = subject
Syntaxeme = no preposition + subjective case
Categorial class = personal

Meaning = object
Syntaxeme = no preposition + accusative case
Categorial class = any

Meaning = causative
Syntaxeme = for + accusative case
Categorial class = attribute

Relation = CAUS
Syntaxeme1 = subject Syntaxeme2 = causative

Linguistic rules

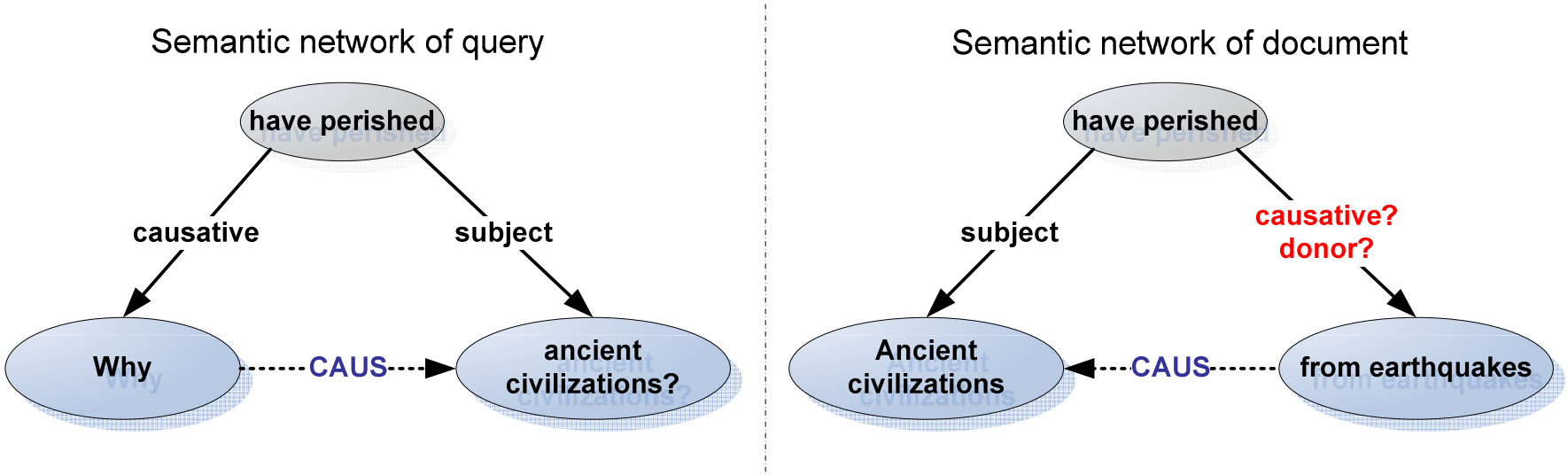
Linguistic rules is context rules for semantic meanings founding.

The rule is a pare <hp, p>, where

hp – antecedent, describes syntaxeme context;

p – consequent, syntaxeme meaning.

Polysemy problem

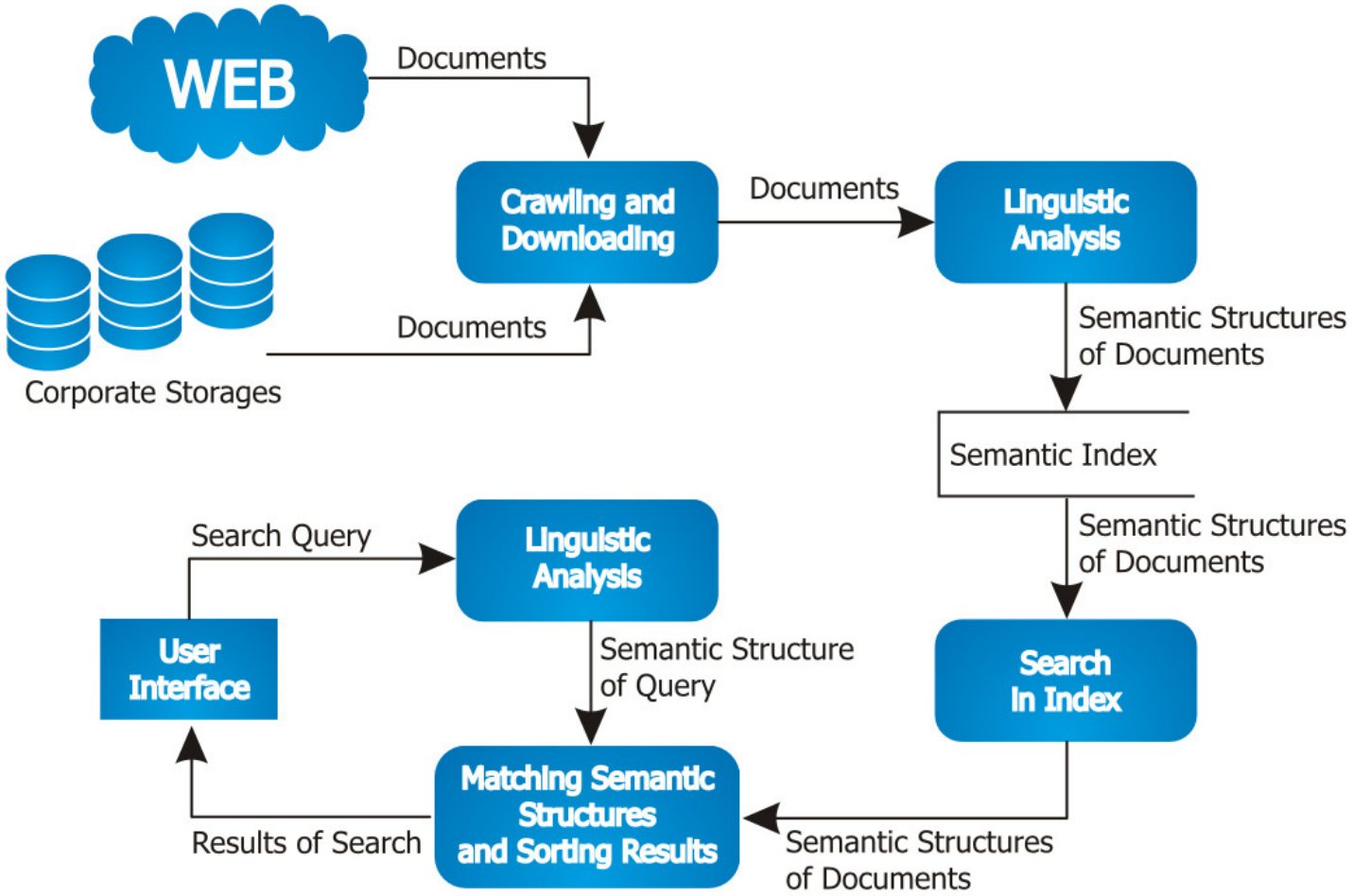


The syntaxeme «from earthquakes» is polysemantic

Rule for polysemy elimination

- If a syntaxeme in *causative case* with preposition *from* is *attributive* and follows a syntaxeme in *nominative* case then the first syntaxeme has the meaning *causative*.

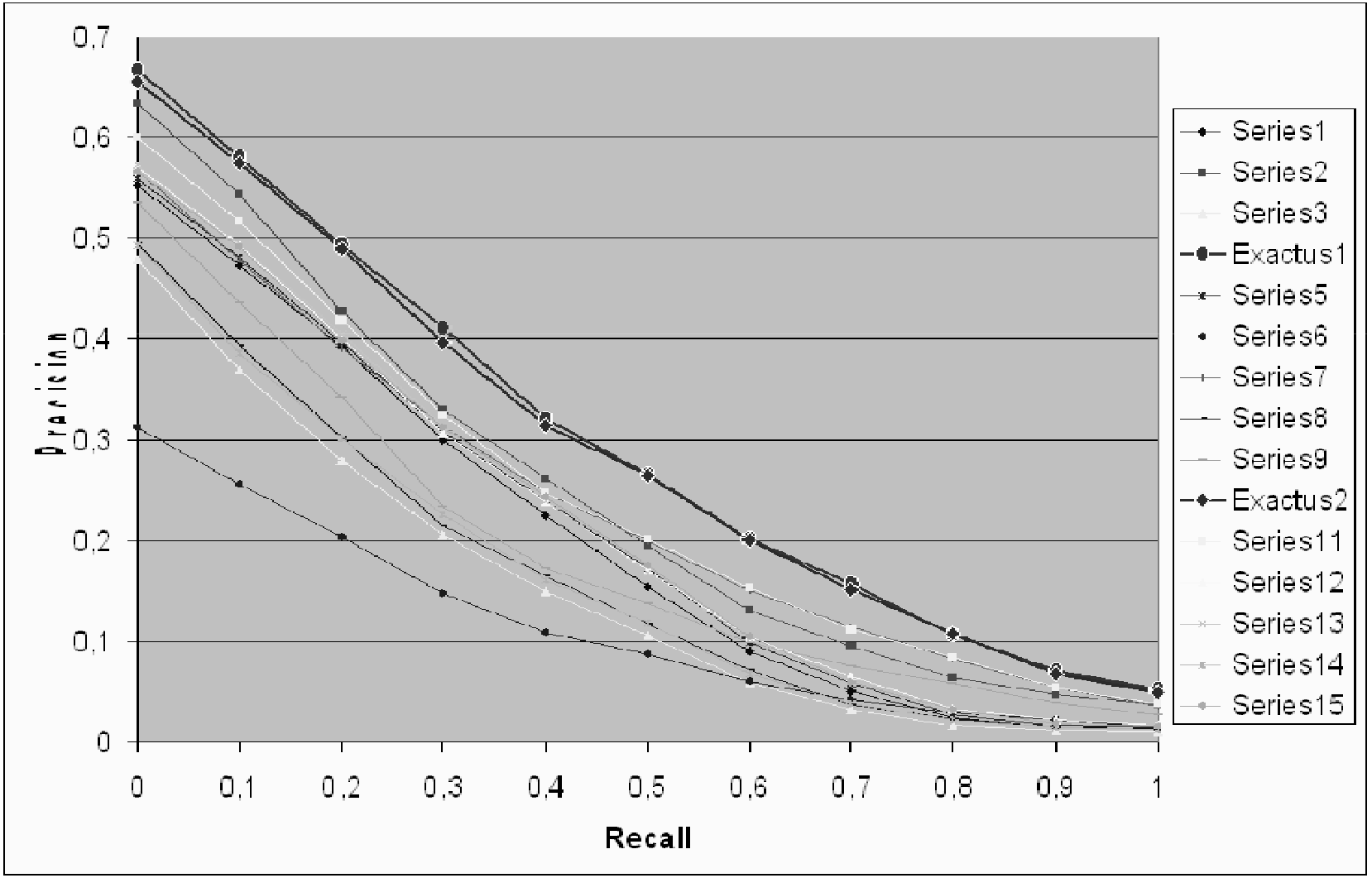
Semantic search engine architecture



Semantic search (4)

- Semantic approach improves precision and recall of the information search and decreases the number of irrelevant documents returned as the result of the search.

Experiments (ROMIP'2008 contest 11-point TREC)



Semantic search (4): advantages

- Support for natural language queries.
- Question answering mode.
- Cross-language and multi-language semantic search.
- Refined methods for morphological, syntactic and semantic analysis.
- Integration of linguistic and statistic search methods.

Contacts

Institute for Systems Analysis
Russian Academy of Science
117312,Moscow,
pr. 60-let Octiabrya, 9
phone/fax: +7 (499) 135-04-63
e-mail: matandra@isa.ru