

Проблемы использования данных из облака LOD для обогащения контента научных баз данных и знаний

З. В. АПАНОВИЧ, А.Г. МАРЧУК

**Институт Систем Информатики имени
А.П. Ершова, СО РАН**

<http://duh.iis.nsk.su/turgunda/Home>

1976	1988	участник	заместитель заведующего кафедрой вычислительной математики ММФ	Новосибирский государственный университет
1979	1988	первое лицо	председатель	Комиссия по системному математическому обеспечению Координационного комитета по Вычислительной технике СССР (КОСМО ККВТ АН СССР)
		участник	участник	Событие "Общим собранием Академии наук избраны п. Сибирск..."
		организатор	организатор	Празднование 10-летия Отдела программирования
		участник		Третий всесоюзный симпозиум "Системное и теоретическое программирование"
1964	1988	участник	заведующий отделом	Институт вычислительной математики и математической геофизики СО РАН
1959	1964	участник	заведующий отделом программирования	Институт математики им. С.Л. Соболева СО РАН

отраж. в документе

Отражение



- The content of the SB RAS Open Archive provides various documents reflecting information about people, research organizations and major events that have taken place in the SB RAS since 1957.
- 20 505 photo documents,
- facts about 10 917 persons
- and 1519 organizations and events.
- The data sets of the Open Archive are available as an RDF triple store, as well as a Virtuoso endpoint.
- Its RDF triple store comprises about 600 000 RDF triples.

- A four-step strategy for the integration of Linked Data into an application consists of:
 - access to linked data,
 - vocabularies (schema, ontology) normalization,
 - identity resolution,
 - data filtering.

[Schultz, A., Matteini, A., Isele, R., Mendes, P.N., Becker, C., Bizer, C.: **How to integrate LINKED DATA into your application.** In: Semantic technology & Business Conference, San Francisco, June 5, 2012. <http://mes-semantics.com/wp-content/uploads/2012/09/Becker-et-al-LDIF-SemTechSanFrancisco.pdf>. (2012)]

- Class
- reflection
- archive-member
- reflection
- collection-member
- location
- person
- photo-doc
- person
- titled
- participation
- email
- photo-doc
- reflection
- phone
- city
- archive-member
- authority
- reflection
- business-activity
- photo-doc
- dating
- reflection
- org-sys
- document
- participation
- org-sys
- phone
- participation
- naming
- person
- opinion
- reflection
- email
- reflection
- collection-member
- collection
- reflection
- archive-member
- location

Query:

```

PREFIX : <DEMO#>

Construct {?o :colleague ?npl.
          ?o :name ?np.
          ?o :works_for ?norg}

where{
  ?s <http://iis.nsk.su/participant> ?o.
  ?s <http://iis.nsk.su/in-org> ?t.
  ?o a <http://iis.nsk.su/person>.
  ?o <http://iis.nsk.su/name> ?np.
  ?t a <http://iis.nsk.su/org-sys>.
  ?t <http://iis.nsk.su/name> ?norg.

  ?sl <http://iis.nsk.su/participant> ?ol.
  ?sl <http://iis.nsk.su/in-org> ?tl.
  ?ol a <http://iis.nsk.su/person>.
  ?ol <http://iis.nsk.su/name> ?npl.
  ?tl a <http://iis.nsk.su/org-sys>.
  ?tl <http://iis.nsk.su/org-classification> ?classif.

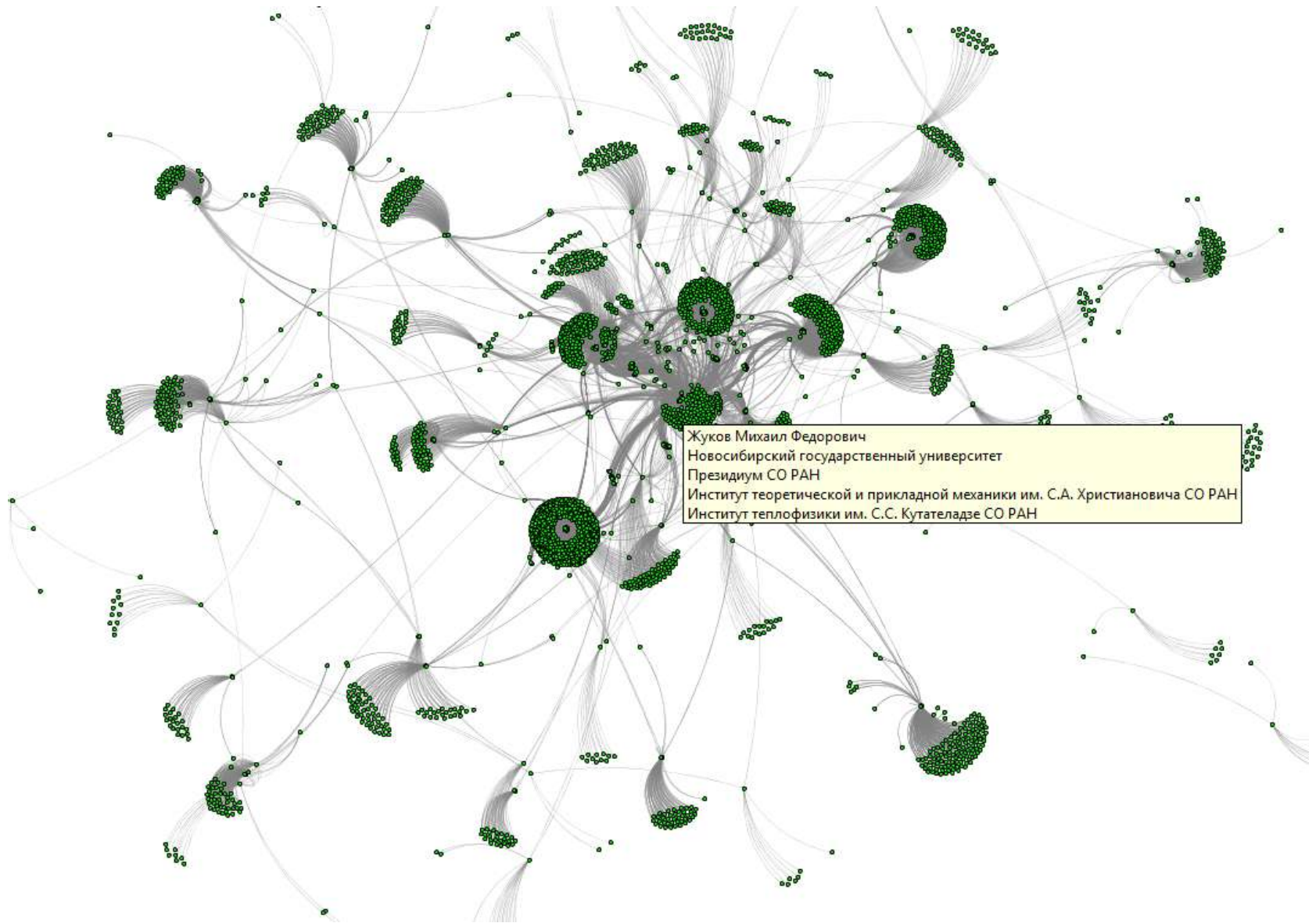
FILTER(?tl=?t && ?ol != ?o && ?classif= "organization" &&langMatches(lang(?npl),"RU"))
} limit 2000

```

Execute clustering

Execute

http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Петров Александр Константинович"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Павловский Евгений Николаевич"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Бубнов Н.Н."@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Чесноков Евгений Николаевич"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Букеев Валерий Александрович"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Корнеев Вадим Давидович"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Наберунов Вадим Исмаилович"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Козаченко Леонид Сергеевич"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Самойлова Елена Ивановна"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Самойленко Владимир Иванович"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Цыганков Вадим Дмитриевич"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Еремеева Мария Владимировна"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Самойленко Владимир Иванович"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Мазур Владимир Григорьевич"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Комаровский Николай Леонид"
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Белкин Николай Михайлович"@ru
http://iis.nsk.su/e20070417_5_15068	<file:///C:/Documents%20and%20Settings/sina/Pa60us8%20cron/vinokurov/15%20june/rdf/DEMO#colleague>	"Толмачев Владимир Артемьевич"



Жуков Михаил Федорович
Новосибирский государственный университет
Президиум СО РАН
Институт теоретической и прикладной механики им. С.А. Христиановича СО РАН
Институт теплофизики им. С.С. Кутателадзе СО РАН

Bone ontology

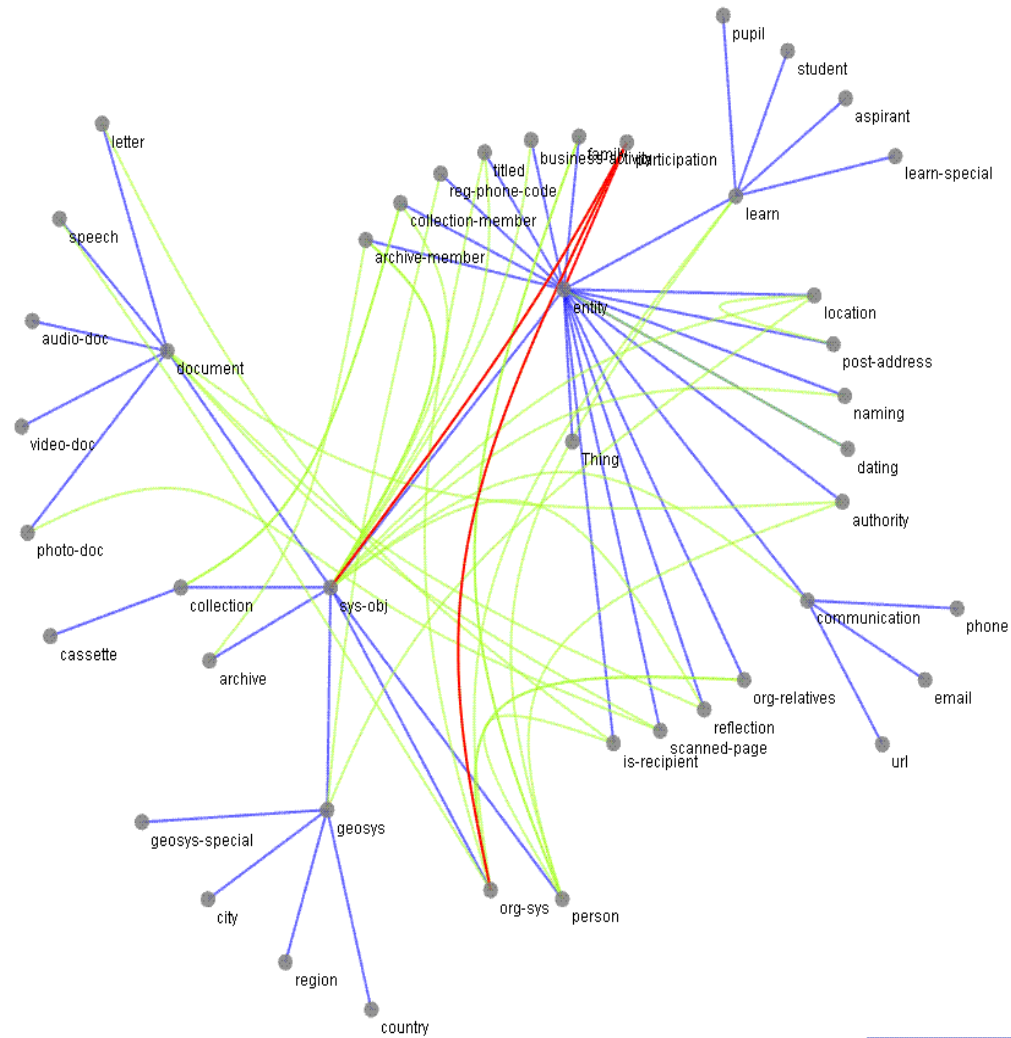
Count of nodes 44

Fix name of node

Show all names

Selected entity class :
participation

Associative relations:
in-org
participant



<http://duh.iis.nsk.su/turgunda/Home>

1976	1988	участник	заместитель заведующего кафедрой вычислительной математики ММФ	Новосибирский государственный университет
1979	1988	первое лицо	председатель	Комиссия по системному математическому обеспечению Координационного комитета по Вычислительной технике СССР (КОСМО ККВТ АН СССР)
		участник	участник	Событие "Общим собранием Академии наук избраны п. Сибирск..."
		организатор	организатор	Празднование 10-летия Отдела программирования
		участник		Третий всесоюзный симпозиум "Системное и теоретическое программирование"
1964	1988	участник	заведующий отделом	Институт вычислительной математики и математической геофизики СО РАН
1959	1964	участник	заведующий отделом программирования	Институт математики им. С.Л. Соболева СО РАН

отраж. в документе

Отражение



Back color

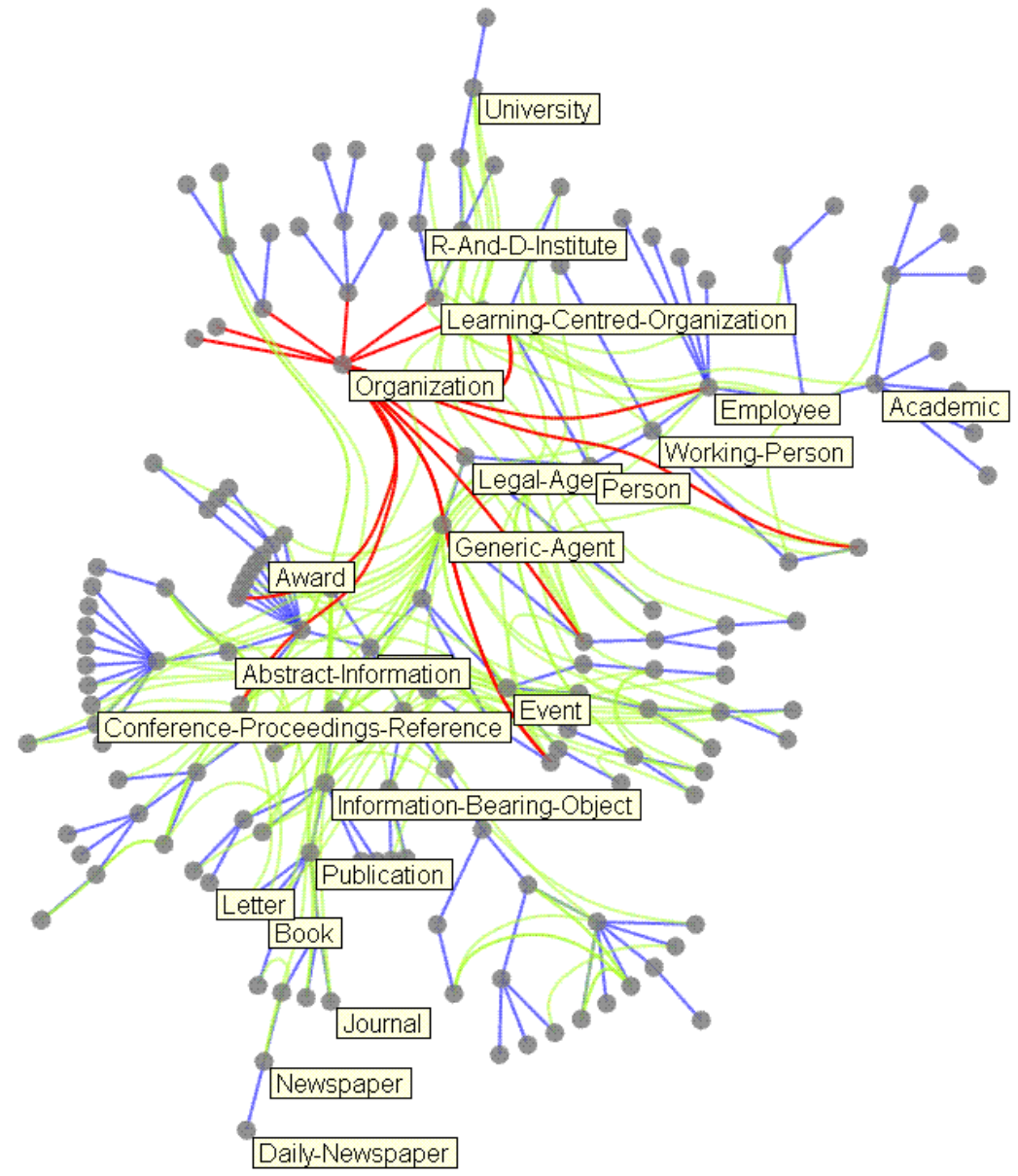
Count of nodes 157

Fix name of node

Show all names

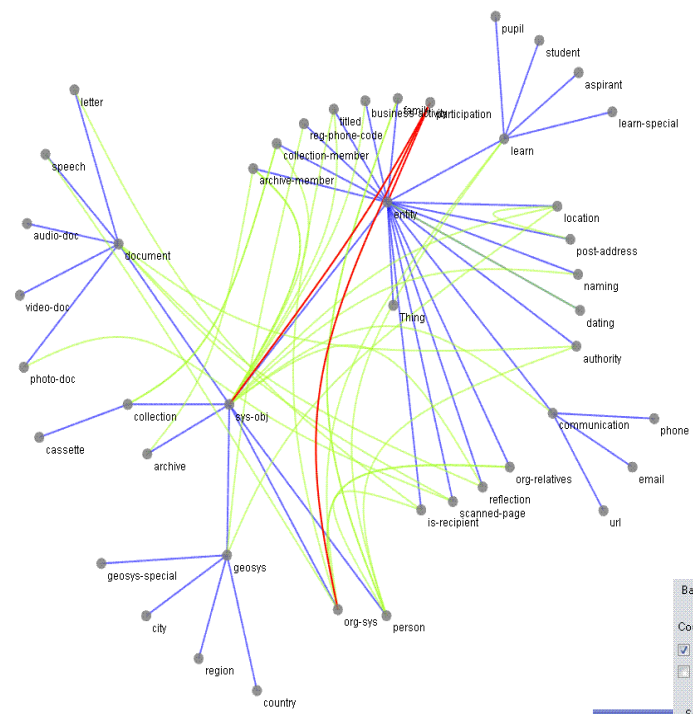
Selected entity class :

- Associative relations:
- affiliated-person
 - funding-source
 - has-affiliation
 - has-size
 - has-sub-unit
 - involves-organization



Selected entity class:
participator

Associative relations:
in-org
participator



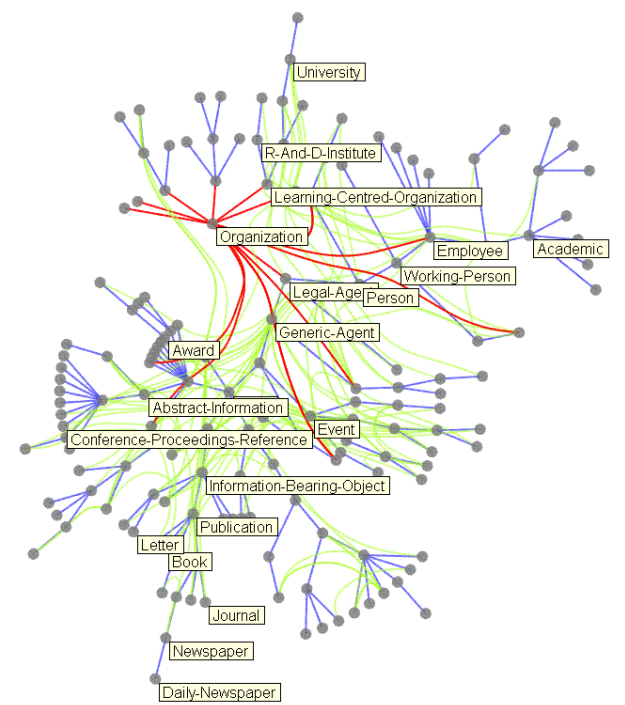
Back color [dropdown]

Count of nodes 157

Fix name of node
 Show all names

Selected entity class:
Organization

Associative relations:
affiliated-person
funding-source
has-affiliation
has-size
has-sub-unit
involves-organization



- It is necessary to establish systematically correspondence between groups of classes and relations of these two ontologies.
- More precisely, a correspondence between one or several groups of the form "**Class1 - relation1 - Class2**" of the AKT Reference ontology and one or several groups of the form "**Class3 - relation2 - Class4 -relation3 - Class5**" of the BONE ontology should be created.
- In particular, a new instance of the **Class4** for every triple \langle **Class1:instance1, relation 1, Class2:instance2** \rangle should be created.

PREFIX iis:<<http://iis.nsk.su#>>

PREFIX akt:<<http://www.aktors.org/ontology/portal#>>

PREFIX akts: <<http://www.aktors.org/ontology/support#>>

CONSTRUCT {

_:p a iis:Class4.

_:p iis:relation2 ?instance1.

_:p iis:relation3 ?instance2.

}

WHERE {

 ?instance1 akt:relation1 ?instance2.

 ?instance1 a akt:Class1.

 ?instance2 a akt:Class2.

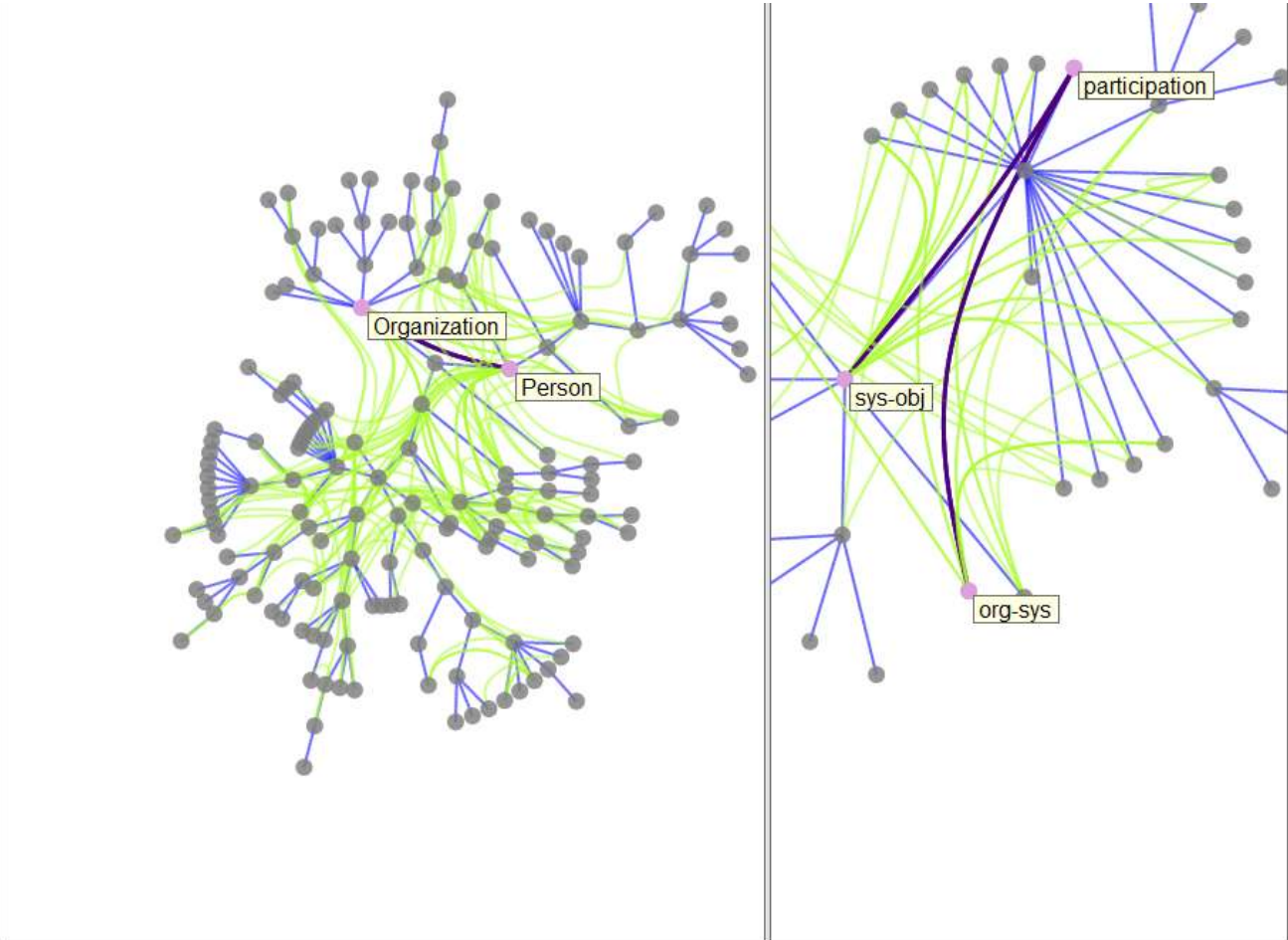
}

View
 Fix name of node
 Show all names

Selected entity class :

Associative relations:

Similar class



Instrument matching

Fix mapping

Person has-affiliation Organization	sys-obj participant participation in-org org-sys	$\langle \text{Person} \rangle \text{ has-affiliation } \langle \text{Organization} \rangle \Rightarrow$ $\langle \text{sys-obj} \rangle \text{ participant } \langle \text{participation} \rangle \text{ in-org } \langle \text{org-sys} \rangle$
---	--	---

- PREFIX iis: <http://iis.nsk.su#>
- PREFIX akt: <http://www.aktors.org/ontology/portal#>
- PREFIX akts: <http://www.aktors.org/ontology/support#>
- PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
- CONSTRUCT {
- _:p a iis:participation.
- _:p iis:in-org ?instance1.
- _:p iis:participant ?instance2.
- }
- WHERE {
- ?instance1 akt:has-affiliation ?instance2.
- ?instance1 a akt:Person.
- ?instance2 a akt:Organization. }

Identity resolution problem

- In the SB RAS Open Archive all persons are specified by means of the «bone: name" attribute.
- The format of this attribute is <Last Name, First Name Middle Name>. This attribute has two options: the Russian-language version and the English-language version. The English version is a transliteration of the Russian version. For example:
- Котов, Вадим Евгеньевич or:
- Kotov, Vadim Evgenievich

- The datasets of RKBExplorer use the *akt:full-name* attribute, and there are many variants of the *akt:full-name* attribute for every instance of the Open Archive. It can be:
- <First Name Last Name>: **Vadim Kotov**
- <First Name First letter of the Middle Name Last Name> **Vadim E.Kotov**
- < First letter of the First Name First letter of the Middle Name Last Name > **V.E. Kotov**,
- **V. Kotov**..., etc.

An example

- It is possible to find at <http://citeseer.rkbexplorer.com>
- 2 persons having full-name attribute **Vadim Kotov**
- **They have different identifiers and different lists of publications**

Vadim Kotov1

<http://citeseer.rkbexplorer.com/id/resource-CSP168322-edfa4d57ca35c11ccbce8e7551242dce>

Duplicate URIs

<http://citeseer.rkbexplorer.com/id/resource-CSP168322-edfa4d57ca35c11ccbce8e7551242dce>
http://kisti.rkbexplorer.com/id/PER_000000000000000146828

Publication

[Control Architecture for Service Grids in a Federation of Utility Data Centers](#)

- **Publications:**

- [Algorithms for Self-Organization and Adaptive Service](#)
- [Control Architecture for Service Grids in a Federation of Utility Data Centers](#)
- [Self-Organizing Control in Planetary-Scale Computing](#)
- [Optimization of E-Service Solutions](#)

- **Organizations:**

- [HP LABORATORIES PALO ALTO](#)
- [PALO ALTO RESEARCH CENTER](#)

- **People**

- [Holger Trinks](#)
- [Artur Andrzejak](#)
- [Sven Graupner](#)

Vadim Kotov2

- <http://citeseer.rkbexplorer.com/id/resource-CSP168328-a3b1b1337798d4fb1e6cbeb53d779d0e>

Duplicate URIs:

<http://acm.rkbexplorer.com/id/person-289779-4255d8bbbc9d678ad18fc77dfb0417d>

<http://citeseer.rkbexplorer.com/id/resource-CSP168328-a3b1b1337798d4fb1e6cbeb53d779d0e>

<http://dblp.rkbexplorer.com/id/people-d32852eb011dfc13e96887308c2f2ca7-0bcd588a7cc1face18d201042b25fb76>

- **People**

- [L. A. Cherkasova](#)

- [Tomas Rokicki](#)

- [Al Davis](#)

- [Ian Robinson](#)

- [Robin Hodgson](#)

- [Gianfranco Ciardo](#)

- **Organizations**

- No results found

- **Publications**

- [Modeling a fibre channel switch with stochastic Petri nets](#)

- [R2: A Damped Adaptive Router Design.](#)

- [Communicating structures for modeling large-scale systems](#)

- [Modeling a scalable high-speed interconnect with stochastic Petri nets](#)

- [Components of congestion control](#)

- [Fibre Channel Fabrics](#)

- **Publications**

Fibre Channel Fabrics: Evaluation and Design.

- The Impact of Message Scheduling on a Packet Switching Interconnect Fabric
- Designing fibre channel fabrics
- Colored Petri Net Methods for Performance Analysis of Scalable High-Speed Interconnects
- R2
- On Net Modeling of Industrial Size Concurrent Systems
- An algebra of concurrent non-deterministic processes
- Concurrent Nondeterministic Processes
- Concurrent Nondeterministic Processes: Adequacy of Structure and Behaviour.
- Descriptive and analytical process algebras
- On Generalized Process Logic
- On structural properties of generalized processes
- Structured Nets

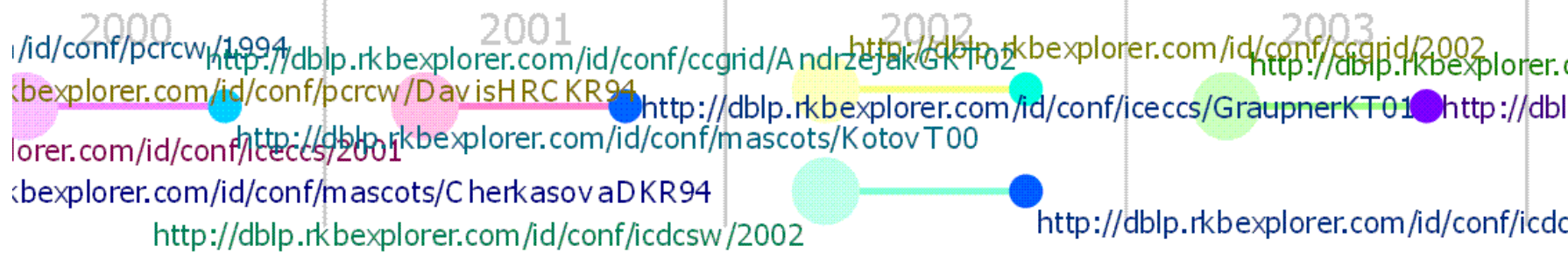
Towards automtical construction of parallel programs

- On the other hand, it is possible to find a list of 32 publications by “**Vadim E. Kotov**” on the <http://dblp.13s.de> address. All these publications belong to “Kotov” from the Open Archive. However, there exists another list of 2 publications belonging to “**Vadim Kotov**”, and only one publication in this list belongs to “Kotov” from the Open Archive.

- 1) Not all publications belonging to the same person are collected together.
- 2) Some publications belonging to different persons are collected together.
- To enrich Open Archive, it is necessary to collect publications from different lists and check if they belong to a person from the Open Archive and not to their homonyms.

Approach1 for identity resolution

- Our experiments with full-text versions of publications have demonstrated that authors usually cite their previous publications.
- This feature allows several people with distinct identifiers to be considered as a single person.
- We have just to explorer self-citation networks!



- After the identification of people from the Open Archive we can add their publications into the Open Archive
- For each instance of the *akt:has-author* **relationship**, it is necessary to generate an instance of the *bone: authorship* **class** along with the *bone: adoc* and *bone: author* relationships linking the instances of the *bone:authorship* class with relevant instances of the *bone:person* and *bone:document* classes.
- **All these transformations can be carried out with a SPARQL- query similar to the one described for the *participation* class.**

Conclusion: current state

The structure of the BONE ontology has been compared to that of the AKT Reference Ontology, and one regular source of their structural difference has been identified.

A template for SPARQL queries that establishes correspondence between groups of classes and relations of the two ontologies has been developed.

A tool generating SPARQL queries on the base of the two ontologies visualization has been implemented

Two new methods of identity resolution are under development

- **Thank you for your attention!**
- **Questions?**