UNIVERSITY OF Southampton
School of Electronics and Computer Science

RKBExplorer.com:

Anatomy of a Semantic Web Application

Hugh Glaser & Ian Millard
4th December 2008, Seoul

Context

- CSAKTiveSpace
  - AKT Project
  - 2003 Semantic Web Challenge winner

- ReSIST - EU Network of Excellence in Resilient Systems
  - Knowledge-enabled infrastructure
  - Jan 2006 – Dec 2008
- KISTI work on Co-reference
  - 2008

UNIVERSITY OF Southampton
School of Electronics and Computer Science

User Interaction

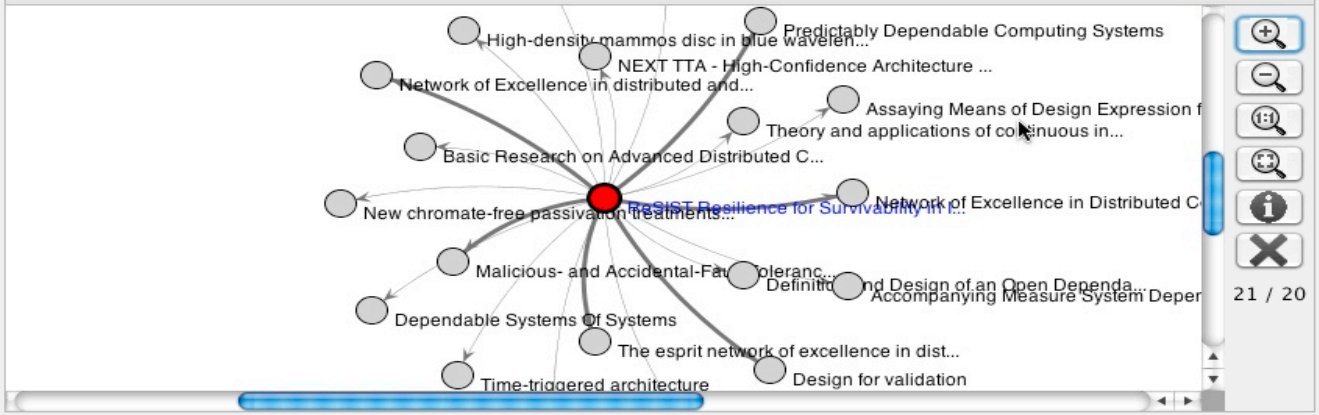
- Semantic MediaWiki
- Custom form interfaces
- Google Maps
- Raw Knowledge Browser

- **RKBExplorer**
- Why do you think that? information

www.rkbexplorer.com/explorer/

Resilience for Survivability in IST

Maximize



Detail

Title:
Resilience for Survivability in IST

Other Names:
ReSIST Resilience for Survivability in IST

Funding:
4500000

Project Leader:
Jean-Claude Laprie

Leading Organisation:
Centre National de la recherche Scientifique

People

- Jean-Claude Laprie
- Aad van Moorsel
- Abdelmajid Khelil
- Achour Mostefaou
- Adnan Noon Mian
- Afraz Jaffri
- Agnan de Bonneval
- Alberto Pasquini

Research Areas

- Information processing, information systems
- Development Failure
- Dependability And Security Analysis
- Dependability And Security Provision

Publications

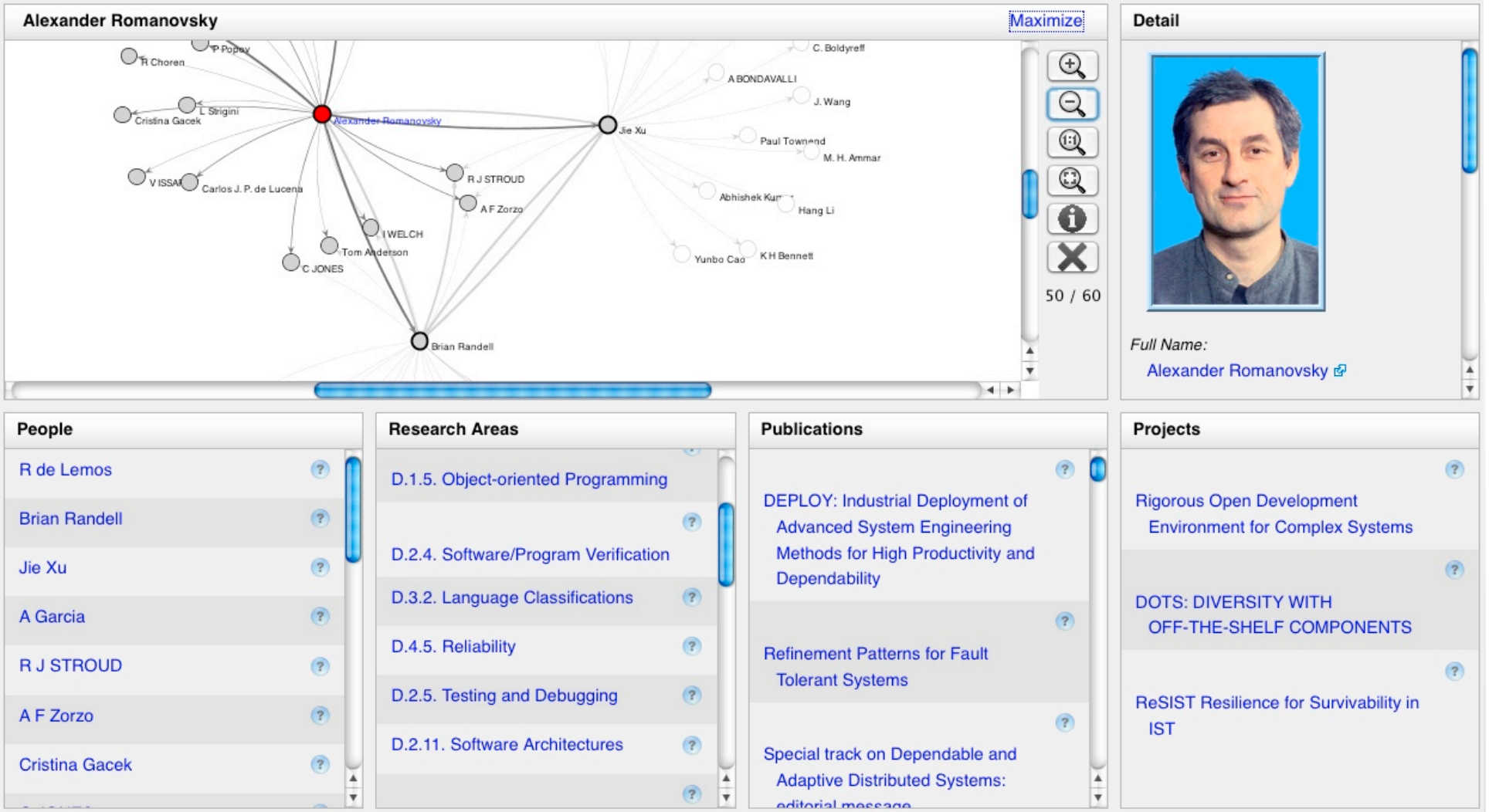
- Revised version in Predictably dependable computing systems
- ASSERT Automated proof based system and software engineering for real-time applications
- Guide de la sûreté de fonctionnement

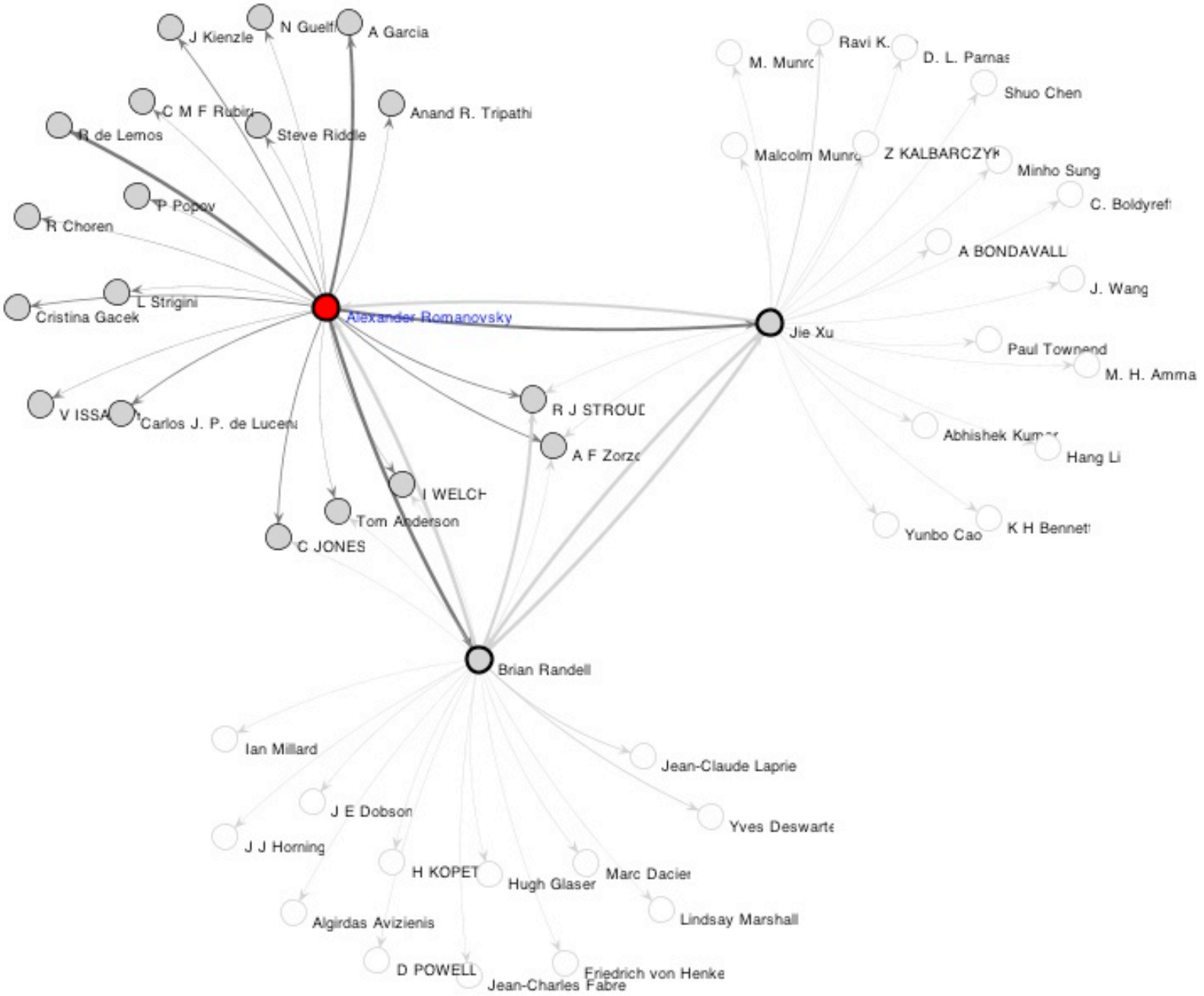
Projects

- Malicious- and Accidental-Fault Tolerance for Internet Applications
- The esprit network of excellence in distributed computing systems architectures
- Design for validation
- Network of Excellence in

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Focusing on a Person




J Kienzle
N Guelfi
A Garcia
C M F Rubiri
Steve Riddle
Anand R. Tripathi
R de Lemos
P Popov
R Choren
L Strigini
Cristina Gacek
Alexander Romanovsky
V ISSA
Carlos J. P. de Lucena
R J STROUD
A F Zorzo
I WELCH
Tom Anderson
C JONES
Brian Randell
Ian Millard
J E Dobson
J J Horning
H KOPET
Hugh Glaser
Marc Dacier
Algirdas Avizienis
Lindsay Marshall
D POWELL
Jean-Charles Fabre
Friedrich von Henke
Jean-Claude Laprie
Yves Deswarte
M. Munro
Ravi K.
D. L. Parnas
Shuo Chen
Malcolm Munro
Z KALBARCZYK
Minho Sung
C. Boldyreff
A BONDAVALLI
J. Wang
Jie Xu
Paul Townend
M. H. Amma
Abhishek Kumar
Hang Li
Yunbo Cao
K H Bennett

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Knowledge Sources

- Partners
- Publications
- Funding Agencies
- Project Wiki
- Courseware
- Resilient-Explicit Computing

- Wide range, don't just look where you expect to find

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Some Underlying Sources

acm.rkbexplorer.com	italy.rkbexplorer.com
budapest.rkbexplorer.com	kaunas.rkbexplorer.com
citeseer.rkbexplorer.com	**kisti.rkbexplorer.com**
cordis.rkbexplorer.com	laas.rkbexplorer.com
courseware.rkbexplorer.com	lisbon.rkbexplorer.com
darmstadt.rkbexplorer.com	newcastle.rkbexplorer.com
dblp.rkbexplorer.com	**nsf.rkbexplorer.com**
deepblue.rkbexplorer.com	pisa.rkbexplorer.com
deploy.rkbexplorer.com	rae2001.rkbexplorer.com
epsrc.rkbexplorer.com	**resex.rkbexplorer.com**
eurecom.rkbexplorer.com	roma.rkbexplorer.com
ft.rkbexplorer.com	southampton.rkbexplorer.com
ibm.rkbexplorer.com	ulm.rkbexplorer.com
ieee.rkbexplorer.com	unlocode.rkbexplorer.com
irit.rkbexplorer.com	wiki.rkbexplorer.com

Range from a few 100 to more than 10,000,000 “facts”

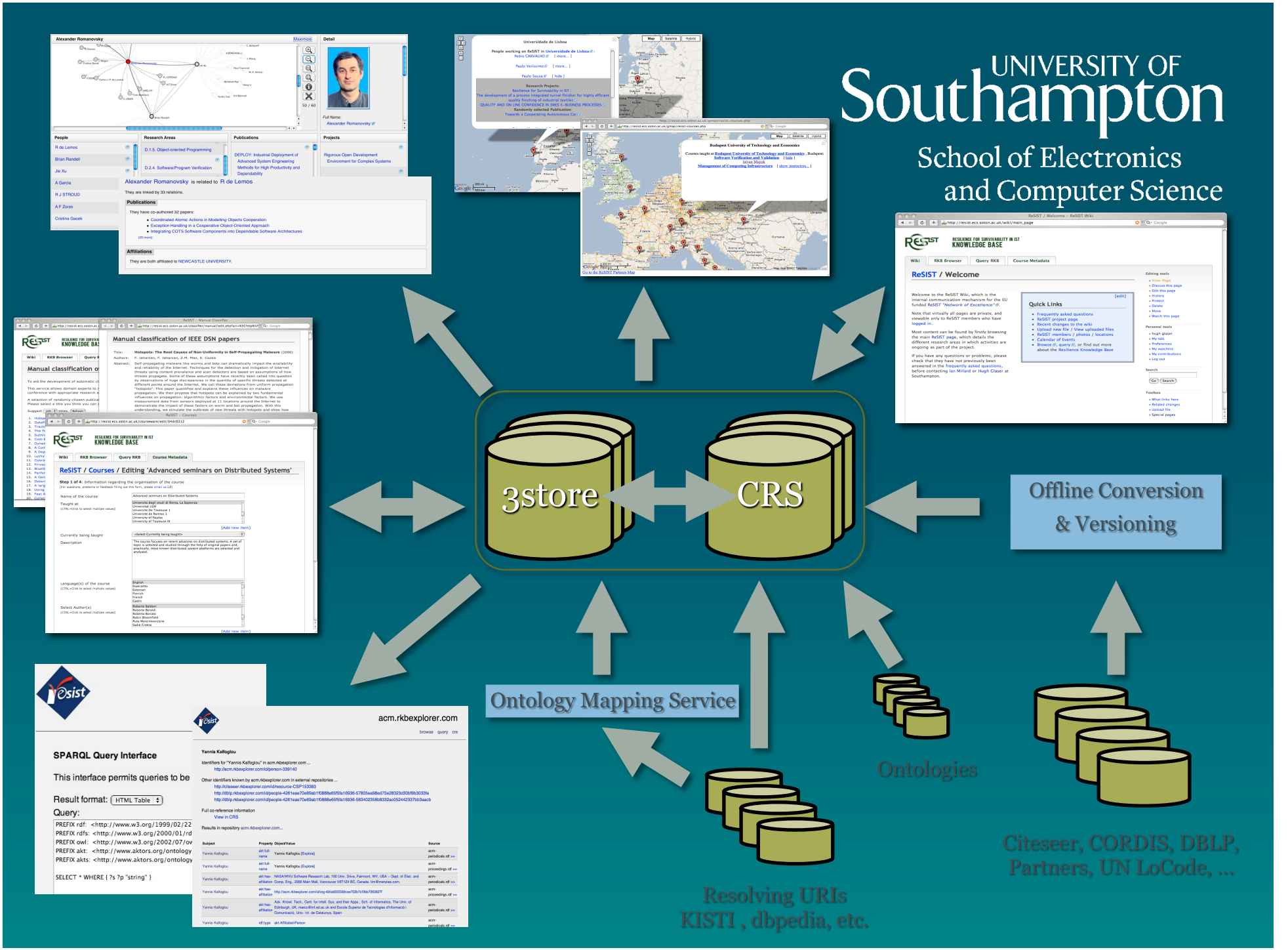
For example

- Statistics for repository kisti.rkbexplorer.com
  - Last data assertion 2008-09-18 17:16:41
  - Number of triples 12815162
  - Number of symbols 3239105
  - Size of RDF dataset 671M

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Semantic Web/Linked Data issues

- The system supports state of the art facilities:
  - Browsing
  - Resolvable URIs
  - SPARQL endpoint
  - CRS (Co-reference knowledge)
  - RDF publishing
  - Semantic Sitemap
  - On a separate domain


UNIVERSITY OF
Southampton
School of Electronics
and Computer Science
3store
CRS
Offline Conversion
& Versioning
Ontology Mapping Service
Ontologies
Resolving URIs
KISTI , dbpedia, etc.
Citeseer, CORDIS, DBLP,
Partners, UN LoCode, ...
Alexander Romanovsky
Alexander Romanovsky is related to R de Lemos
They are linked by 33 relations.
Publications
Affiliations
They are both affiliated to NEWCASTLE UNIVERSITY.
Manual classification of IEEE DSN papers
ReSIST / Courses / Editing 'Advanced seminars on Distributed Systems'
ReSIST / Welcome
Quick Links
SPARQL Query Interface
This interface permits queries to be
Result format:
HTML Table
Query:
acm.rkbexplorer.com
Yannis Kalfoglou

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Communication

- Ontologies
  - General Scientific Endeavour
  - Domain-specific
  - Support (geospatial, etc)
- Open Local Knowledge – HTTP
  - Resolvable URIs
  - SPARQL
- Uses Remote Knowledge
  - Resolves URIs with caching

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Components 1

- Semantic Web infrastructure throughout
- Triplestore for each source
  - Putting the Web in Semantic Web
  - Stores RDF – (Subject, Predicate, Object)
  - We use 3store
- Linked Data
  - 303 and content negotiation architecture with caching

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Components 2

- Co-Reference Subsystem
  - CRS – more later
- Community of Practice Analysis
  - Why do you think that?
- Ontology Mapping
  - Dealing with other Ontologies
- NLP for text classification
- Caching everywhere

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Components 3

- Application Middleware
  - URI Equivalence Closure
  - RDF Graph Closure
- Semantic Sitemap
  - Facilitate Search Engines

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Co-Reference

- Co-Reference is a Big Problem
  - Identifying multiple URIs for one resource
  - Rejecting incorrectly conflated resources
  - Publishing
  - Using
- Coldstart
  - A serious problem
  - Nothing is linked to anything

Co-Reference Closure

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Complete Co-Reference Information

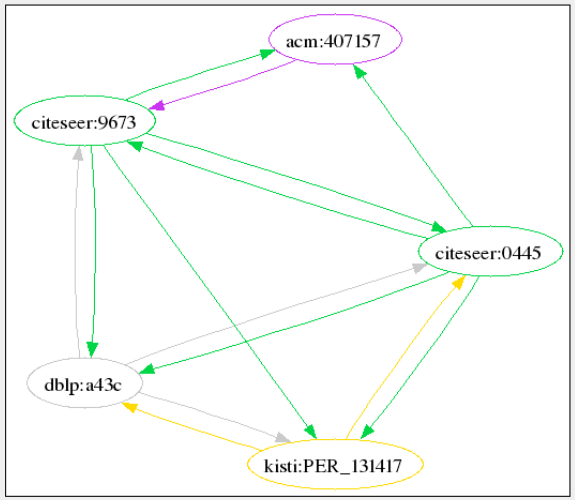
This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevent CRS knowledge bases.

<http://kisti.rkbexplorer.com/id/PER_0000000000000131417> Go

Equivalent URIs...

1. **(Canon)** http://acm.rkbexplorer.com/id/person-407157
2. http://citeseer.rkbexplorer.com/id/resource-CSP179673
3. http://citeseer.rkbexplorer.com/id/resource-CSP180445
4. http://dblp.rkbexplorer.com/id/people-1ec5a600299222dd6374695ef5214f05-90d423eb148125a6e5c573dc5a15a43c
5. http://kisti.rkbexplorer.com/id/PER_0000000000000131417

The following diagram shows the interconnectivity between the CRS knowledge bases which maintain the context-dependent representation of coreference for each of the RKBExplorer domains.



Seungwoo Lee

Showing information queried from all repositories ...

Showing information queried from all repositories ...

Subject	Property	Object/Value	Source
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-periodicals.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-proceedings.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	dblp-publications-2001.rdf >>
Seungwoo Lee	akt:has-affiliation	Electrical and Computer Engineering Division, Pohang University of Science & Technology (POSTECH), Pohang, South Korea.gblee@postech.ac.kr	acm-periodicals.rdf >>
Seungwoo Lee	akt:has-affiliation	POSTECH, Pohang, Korea	acm-proceedings.rdf >>
Seungwoo Lee	kisti:engNameOfPerson	Seungwoo Lee [Explore]	datatypeproperties.ttl >>
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	akt:Person	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	PER_char(20)*^^	datatypeproperties.ttl >>
Seungwoo Lee	rdf:type	PER_char(20)*^^	objectproperties.ttl >>
Seungwoo Lee	rdf:type	PER_char(20)*^^	resources.ttl >>

Subject	Property	Object	Source
Automatic acquisition of named entity tagged corpus from world wide web	akt:has-author	Seungwoo Lee	acm-proceedings.rdf >>
A Corpus-Based Learning Method of Compound Noun Indexing Rules for Korean	akt:has-author	Seungwoo Lee	acm-periodicals.rdf >>
SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP.	akt:has-author	Seungwoo Lee	dblp-publications-2001.rdf >>
A Corpus-Based Learning Method of Compound Noun			dblp-publications-

UNIVERSITY OF Southampton
School of Electronics and Computer Science

CRS – Consistent Reference Service

- A service to manage and publish co-referent information
- Identify co-referent pairs using a set of tools
- Assert into the CRS
- Query the CRS
  - URI_i -> { URI_1, ..., URI_i, ..., URI_n }
- Recommend a Canon

UNIVERSITY OF Southampton
School of Electronics and Computer Science

CRS continued

- CRS Policies are defined by context
  - Often one per Triplestore
  - Can be many per Triplestore for different purposes
  - May not be associated with a particular Triplestore
- Maintenance
  - Provenance
  - Rollback
- Can be used to infer owl:sameAs

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Dealing With Non-SPARQL KBs

- The RKBExplorer application uses SPARQL to query the KBs
  - But needs to access data from KBs that only offer resolvable URIs
- So resolve such a URI
- Cache the RDF with associated resolved RDF locally
- Query the local cache

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Dealing With Different Ontologies

- The RKBExplorer application uses a particular ontology
  - Some KBs will use different ontologies
  - Eg kisti.rkbexplorer.com
- One solution
  - Represent the ontology relationship in RDF (as far as possible)
  - Resolve the URI through the mapping service to get RDF in the required ontology

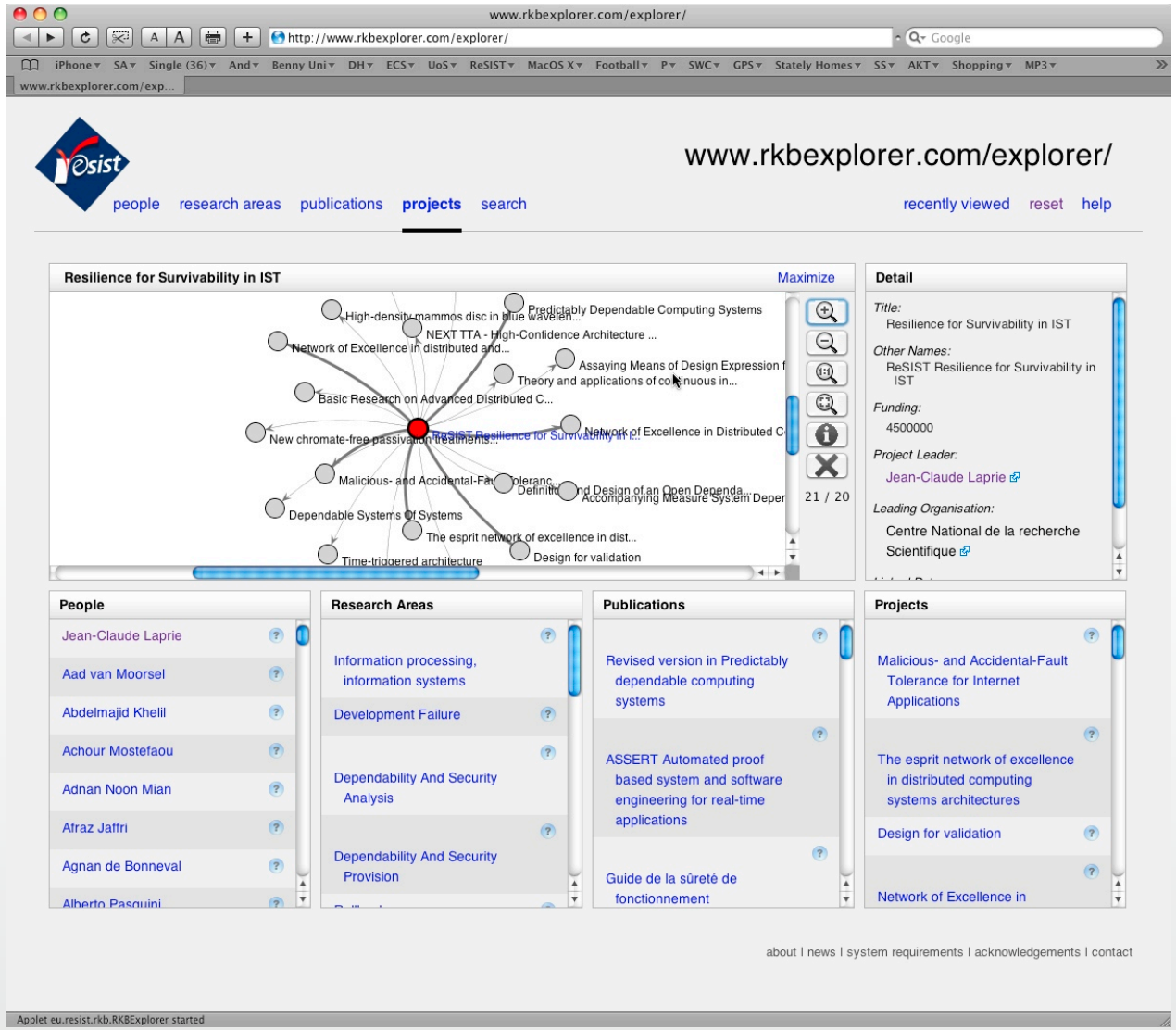
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Concluding Remarks

- Major Data Fusion using Semantic Web Technologies
- Many things can be cast in a Semantic Web framework
- Linked Data works pretty well
- RDF works pretty well
- A little Ontology goes a long way
- Co-Reference is the biggest problem
  - But is tractable

RKBExplorer.com/explorer/ – Try it!

UNIVERSITY OF Southampton
School of Electronics and Computer Science



http://eprints.ecs.soton.ac.uk/16946