



HAL
open science

Interlinking two institutional KOS about Agroecology: using LOD Agrovoc to circumvent the language barrier in identifying terminological intersections

Sophie Aubin, Pascal Aventurier, Ivo Júnior Pierozzi, Leandro Henrique
Mendonça Oliveira

► To cite this version:

Sophie Aubin, Pascal Aventurier, Ivo Júnior Pierozzi, Leandro Henrique Mendonça Oliveira. Interlinking two institutional KOS about Agroecology: using LOD Agrovoc to circumvent the language barrier in identifying terminological intersections. DC2015 International Conference on Dublin Core and Metadata Applications, Sep 2015, São Paulo, Brazil. DC2015 International Conference on Dublin Core and Metadata Applications, 3 p., 2015. hal-01192923

HAL Id: hal-01192923

<https://hal.science/hal-01192923v1>

Submitted on 19 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interlinking two institutional KOS about Agroecology: using LOD Agrovoc to circumvent the language barrier in identifying terminological intersections

Sophie Aubin

sophie.aubin@versailles.inra.fr
INRA, France

Pascal Aventurier

pascal.aventurier@avignon.inra.fr
INRA, France

Ivo Pierozzi Júnior

ivo.pierozzi@embrapa.br
Embrapa, Brazil

Leandro H. M. Oliveira

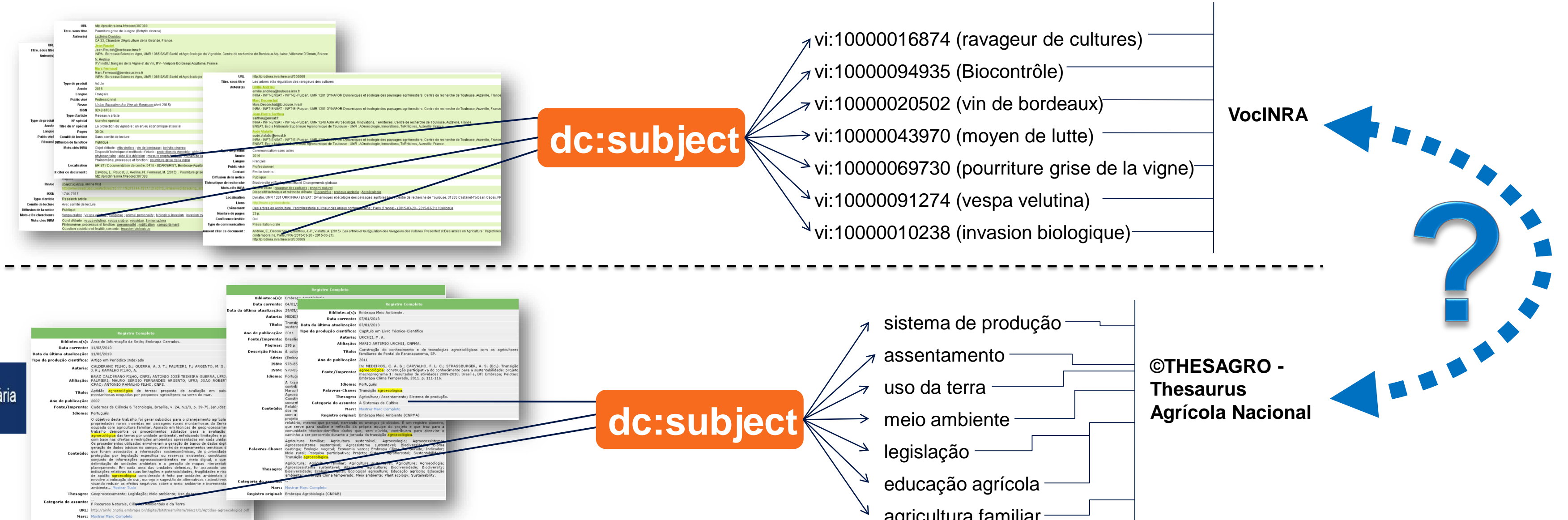
leandro.oliveira@embrapa.br
Embrapa, Brazil

Context and aims of the work



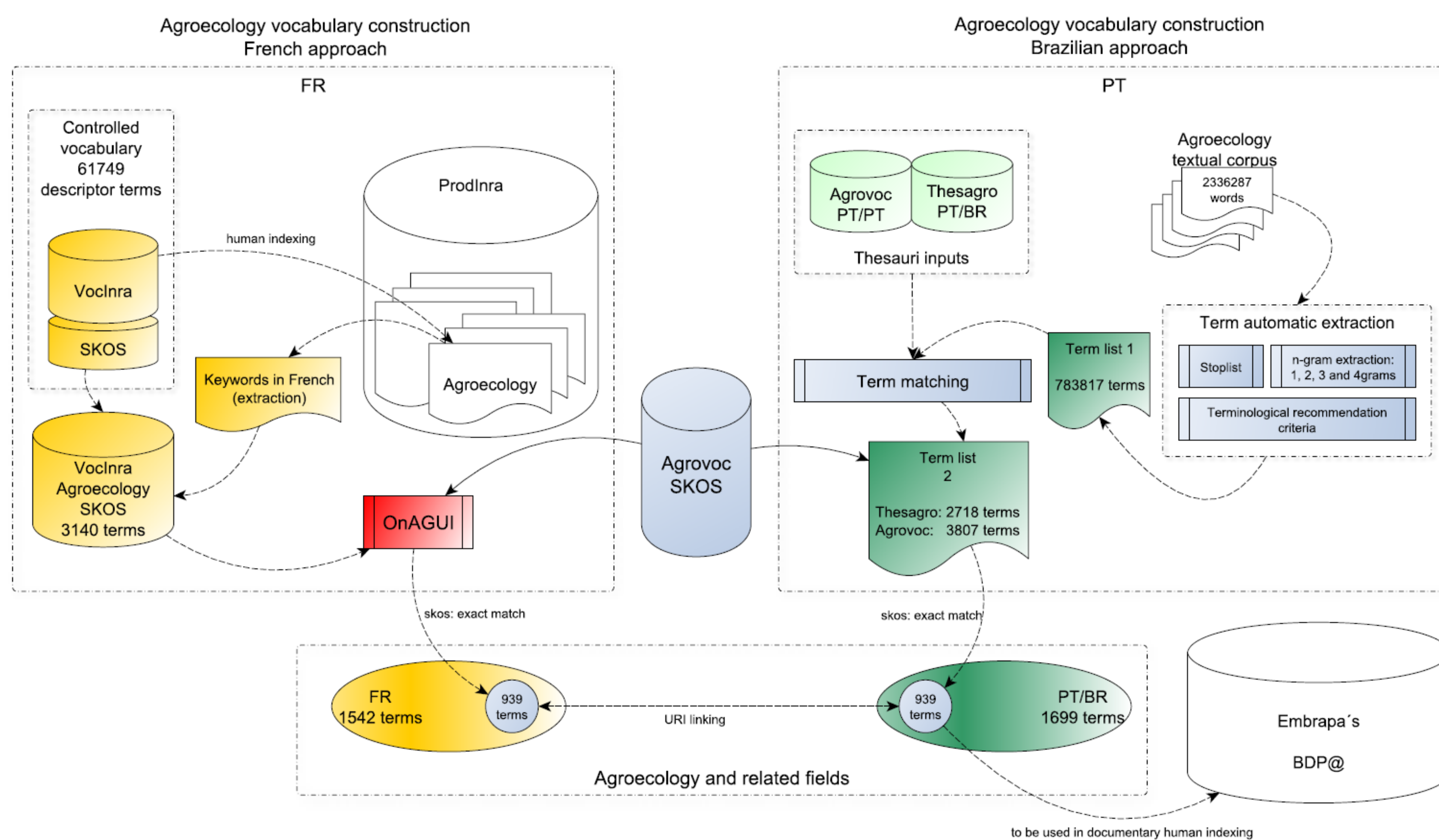
Agroecology

- a common domain of high interest
- the need to consolidate dedicated vocabularies
- the need to compare and share knowledge



Material and methods

General Methodological Approach



Agrovoc Linked Open Data (Agrovoc LOD) serialized in RDF SKOS and offering concept labels in French and Portuguese was chosen as the key solution..

Knowledge Organization System subsets:

- INRA: 3,140 French terms from VocINRA used to manually index 2,145 publications about Agroecology in the institutional repository ProDINRA;
- Embrapa : semi-automatic term extraction from 260 full papers about Agroecology and a term matching using a specific tool (Pierozzi Júnior *et al.*, 2014) towards Thesagro (a Brazilian Portuguese thesaurus) and Agrovoc-PT.

Alignment

- Onagui (Mazuel and Charlet, 2010), an open source tool designed to help alignment of vocabularies in SKOS or OWL, was used to align the VocINRA concepts with those in Agrovoc.
- The exact match relation between Thesagro and Agrovoc-PT and was used to build the final Embrapa term subset.

Results

Examples of exact matched comum terms for both French (VocINRA) and Brazilian Portuguese (EmbrapaVoc) Vocabularies

Agrovoc URI	VocINRA	EmbrapaVoc
agrovoc:c_6701, "rural development"	vi:1000006218, "développement rural"	"desenvolvimento rural"
agrovoc:c_7645, "technology transfer"	vi:10000019611, "transfert de technologie"	"transferência de tecnologia"
agrovoc:c_7973, "tropical climate"	vi:1000004499, "climat tropical"	"clima tropical"
agrovoc:c_10195, "cucumbers"	vi:1000004499, "concombre"	"pepino"
agrovoc:c_2746, "ewes"	vi:1000002777, "brebis"	"ovelha"

Examples of exact matched terms for French Vocabulary (VocINRA)

Agrovoc URI	VocINRA	EmbrapaVoc
agrovoc:c_14187, "selective grazing"	vi:10000047526, "système de pâturage"	--
agrovoc:c_1394, "cauliflowers"	vi:1000004244, "chou fleur"	--
agrovoc:c_16096, "agrosilvopastoral systems"	vi:10000018929, "système agrosilvopastoral"	--
agrovoc:c_2595, "environmental control"	vi:10000089386, "protection de l'environnement"	--

Examples of exact matched terms for Brazilian Portuguese Vocabulary (EmbrapaVoc)

Agrovoc URI	VocINRA	EmbrapaVoc
agrovoc:c_8086, "urban population"	---	"população urbana"
agrovoc:c_4889, "molasses"	---	"melaço"
agrovoc:c_15521, "social indicators"	---	"indicador social"
agrovoc:c_13127, "passion fruits"	---	"maracujá"

prefix agrovoc=http://aims.fao.org/aos/agrovoc/

prefix vi=http://opendata.inra.fr/resources/vocinra#

Conclusions

- Agrovoc LOD allows
 - (1) the identification of common terms used for the two institutions and
 - (2) a set of specific terms in each language that might be incorporated into each other's vocabularies;
- This work also highlights some difficulties in the translation of certain terms in Agrovoc which will improve this multilingual vocabulary;
- The alignment between the INRA and Embrapa vocabularies prepares these two vocabularies to be linked when published in LOD;
- This conciliatory methodological model can be strengthened and systematized so that Embrapa and INRA can consider their contribution to broader initiatives in the agricultural domain like the project for a Global Agricultural Concept Scheme (Baker and Suominen, 2014).

References

- Agrovoc LOD: <http://aims.fao.org/standards/agrovoc/linked-open-data>.
- Baker, Thomas and O. Suominen. Global Agricultural Concept Scheme (GACS): A multilingual thesaurus hub for Linked Data. 2014 http://aims.fao.org/sites/default/files/posts/attachments/GACS_Integration_Proposal_1.0_3.pdf.
- Mazuel Laurent and Jean Charlet "SPIM-AlignmentGUI - un logiciel d'aide à la réalisation d'alignements entre ontologies 2009. Inria http://ic2009.inra.fr/docs/posters/MazuelCharlet_Poster_IC2009.pdf.
- Pierozzi, Ivo Júnior, Marcia Izabel Fugisawa Souza, Tércia Zavaglia Torres, Leandro Henrique Mendonça de Oliveira and Leonardo Ribeiro Queiros. Gestão da informação e do conhecimento. In: Tecnologias da informação e comunicação e suas relações com a agricultura. Brasília, DF: Embrapa, 2014. Cap. 12. p. 237-260. URL: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/119627/1/capitulo12-085-14.pdf>.
- OnAGUI - Ontology Alignment GUI : <http://sourceforge.net/projects/onagui/>.
- Stoilos Giorgos;Stamou, Giorgos; Kollias, Stefanos (2005). A String Metric for Ontology Alignment. The Semantic Web – ISWC 2005. Lecture Notes in Computer Science Volume 3729, pp 624-637.