



Working in Tech-mining . Current developments in the Basque Country

- ☐ **MAPPING THE SCIENCE OF WASTE RECYCLING
Evolution of Research From 2002 to 2012**
- ☐ **Patent Overlay Maps. Spain and Basque Country**
- ☐ **Patent Analysis to Create New Technology Based Firms**



MAPPING THE SCIENCE OF WASTE RECYCLING

Evolution of Research From 2002 to 2012

Index

- 1st Step - Setting the target.
- 2nd Step - Choosing databases.
- 3rd Step - Downloading the data.
- 4th Step - Data import and merging.
- 5th Step - Cleaning the data.
- 6th Step - Generating co-occurrence matrix.
- 7th Step - Visualizations.



1st Step

Setting the target.

This study will use bibliometric databases to map the research taking place around waste recycling, and the evolution from year 2002 to year 2012 will be analyzed.

A versatile boolean approach is designed for «capturing» this research from multiple databases.

Vantage Point text mining will be used for

- Merging the items retrieved from several databases.**
- Cleaning the duplicities.**
- Cleaning the keywords indexed in «author keyword» field, building a thesaurus in this process.**
- Building a keyword co-occurrence matrix.**



2nd Step

Choosing databases.

University of Connecticut database locator (University of Connecticut 2012), for finding environmental sciences specialized data sources.

EBSCO Green File was selected as specialized database.

SCOPUS and SCI were selected as generalistic, wide-coverage databases.

SSCI database was included given the relevant role played by social science in waste recycling field, as detected in previous works (Garechana et al. 2012b).

3rd Step

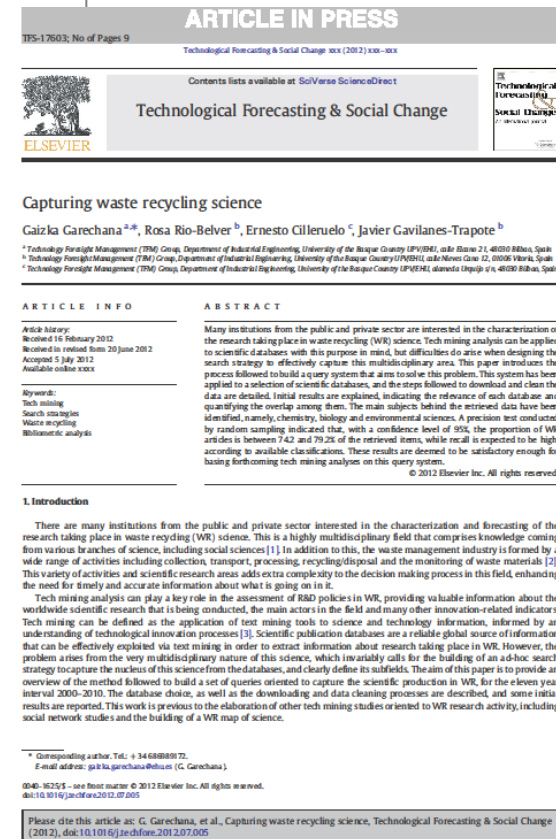
A versatile, flexible boolean query approach was the choice to get the information contained in several databases. The query system required slight adaptations to the syntax of each particular database.

This system consists of 32 queries complemented by an optional query and a exclusion query aimed at eliminating noise from retrieved items.

Really, the extraction of the journal articles corresponding to Waste Recycling science has been a matter of research by itself.

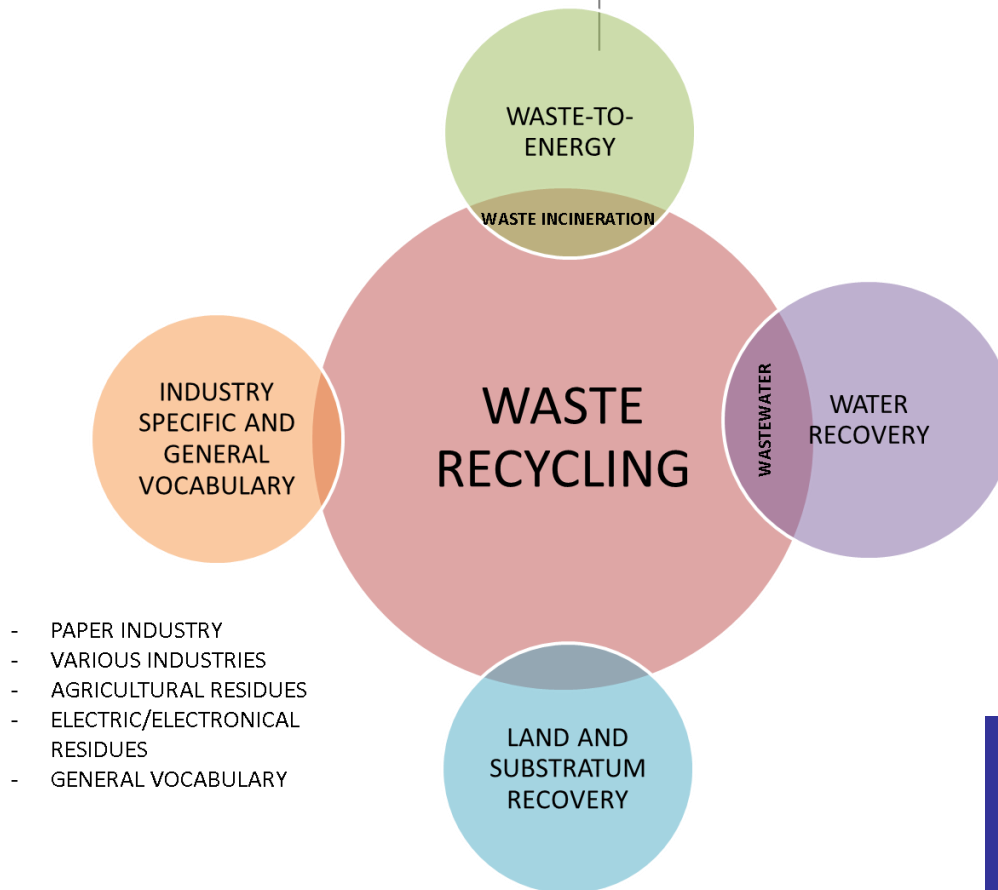
This search strategy has been approved by experts on the field.

Downloading the data.



3rd Step Downloading the data.

Venn diagram reflecting the main areas detected in waste recycling previous characterization, and some overlap zones.



Tracking the evolution of waste recycling research using overlay maps of science
Gaizka Garechana^{a,*}, Rosa Rio^b, Ernesto Cilleruelo^c, Javier Gavilanes^b

^a Research, Technology and Management (RTM) Group, Department of Industrial Engineering, University of the Basque Country (UPV/EHU), Leizaola 21, 48930 Billbao, Spain
^b Research, Technology and Management (RTM) Group, Department of Industrial Engineering, University of the Basque Country (UPV/EHU), Leizaola 21, 48930 Billbao, Spain
^c Research, Technology and Management (RTM) Group, Department of Industrial Engineering, University of the Basque Country (UPV/EHU), Leizaola 21, 48930 Billbao, Spain

ARTICLE INFO

Article history:
Received 22 July 2011
Accepted 20 January 2012
Available online 18 February 2012

Keywords:
Tech mining
Science map
Research forecasting
Recycling trends

ABSTRACT

Tracking the evolution of research in waste recycling science (WRS) can be valuable for environmental agencies, as well as for recycling businesses. Maps of science are visual, easily readable representations of the cognitive structure of a branch of science, a particular area of research or the global spectrum of scientific production. They are generally built upon evidence collected from reliable sources of information, such as patent and scientific publication databases. This study uses the methodology developed by Rafols et al. (2010) to make a "fuzzy overlay map" of WRS upon a basemap reflecting the cognitive structure of all journal-published science, for the years 2005 and 2010. The analysis has taken into account the cognitive areas where WRS articles are published and the areas from where it takes its intellectual nourishment, paying special attention to the growing trends of the key areas. Interpretation of results lead to the conclusion that extraction of energy from waste will probably be an important research topic in the future, along with developments in general chemistry and chemical engineering oriented to the recovery of valuable materials from waste. Agricultural and material sciences, together with the combined economics, politics and geography field, are areas with which WRS shows a relevant and even increasing cognitive relationship.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Changes in the amount and composition of waste, in prices of raw materials, and in the performance of recycling processes, are some of the facts that condition the research trends in waste recycling science (WRS). Tracking the evolution of these research activities is an issue of importance for environmental agencies and for recycling business, worldwide. Tracking allows decision makers to develop R&D strategy with empirical facts instead of bare intuition. Tech mining tools can be integrated in a decision support system to reach this goal (Porter and Newman, 2011), making it possible to analyze the very core of worldwide scientific production: the patent and scientific publication databases. Tech mining consists of the application of text mining tools to science and technology information databases, looking for answers to a wide set of key technology management questions (Porter and Cunningham, 2005). Some applications of tech mining include tracking the specialization of different countries in a specific scientific development (Islam and Miyazaki, 2010), or anticipating emerging research fronts, by analysis of citations (Shibata et al., 2008).

Maps of science are a feasible using tech-mining tools; these maps are visual representations of the cognitive structure of a branch of science, a particular area of research or the global spectrum of scientific production. The strong points of these visualizations are their readability and the reliability of the sources of information upon which they are built (Noyons, 2005).

This paper aims to analyze the evolution of WRS following the overlay method introduced by Rafols et al. (2010), a method successfully employed with different basemaps to analyze the degree of interdisciplinarity of scientific production in (Rafols et al., 2011; Porter and Rafols, 2009; Rafols and Meyer, 2010) and the cognitive diffusion of particular research areas (Leydesdorff and Rafols, 2011; Kiss et al., 2010).

The overlay maps could also be used to locate bodies of research in the global map of science, obtaining an attractive and intuitive view of the research profile of a university, laboratory, country or enterprise. This application can be particularly useful for benchmarking activities. Some interesting examples can be found in the ISI website (Porter et al., 2010 updated).

The scientific publications pertaining to WRS, as contained in ISI's Web of Science (WOS), have been processed using the text mining tool Vantage Point (Search Technology, 2011 updated) for years 2005 and 2010, and the results overlaid upon the global map of science as shown in Rafols et al. (2010). This global map of science is a picture of the cognitive structure of all science, where the different cognitive areas are represented by Subject Categories

* Corresponding author. Tel.: +34 946088972.
E-mail addresses: gaizka.garechana@ehu.es (G. Garechana), rosario.rio@ehu.es (R. Rio), ernesto.cilleruelo@ehu.es (E. Cilleruelo), javier.gavilanes@ehu.es (J. Gavilanes).

0950-4230/\$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wasman.2012.01.017

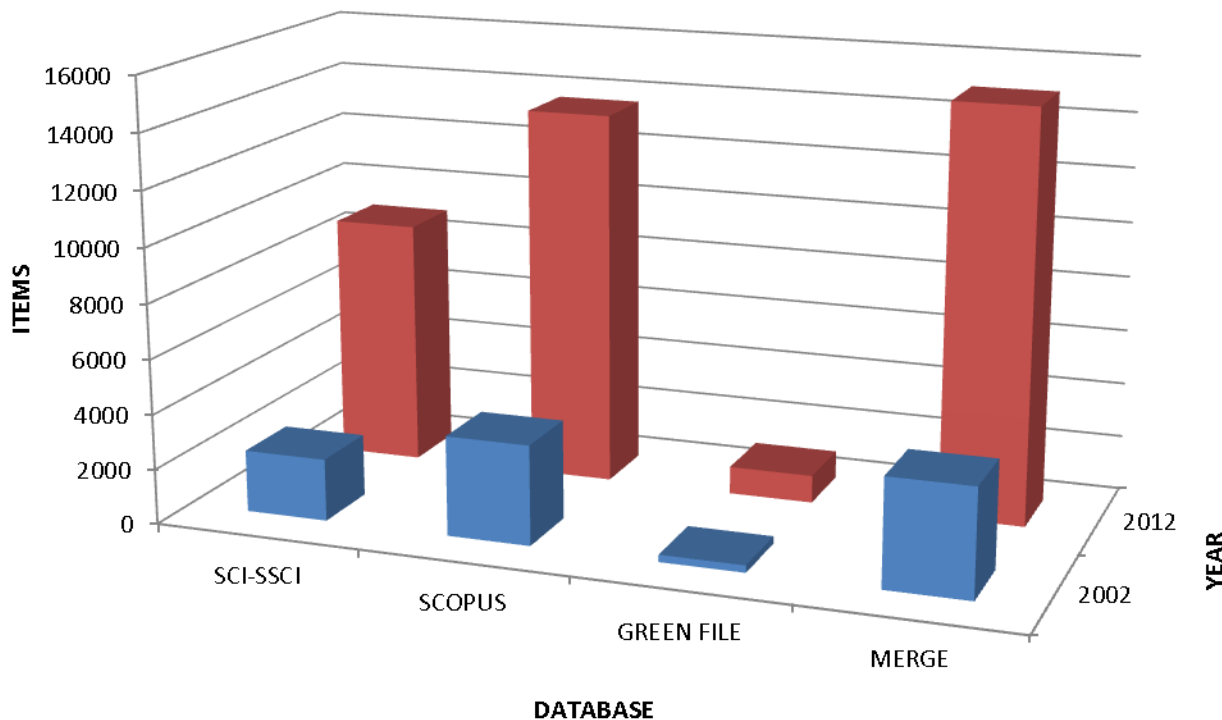
We adopt an inclusive definition approved by European Environmental Agency

« A method of recovering waste as resources which include the collection and often involving the threatment of waste products for use as a replacement of all or part of the raw material in a manufacturing process »

3rd Step Downloading the data

Journal articles
retrieved and
database overlap

- Number of journal articles retrieved in the databases for years 2002. The column «merge» points out a significant overlap among databases. Number of publications increases notably from 2002 to 2012.





4th Step

Data import and merging.

Files were downloaded from SCOPUS, WOS and Green File in proper formats for importing them using filters available at <https://thevantagepoint.com/> and other customized filters made by VP support team.

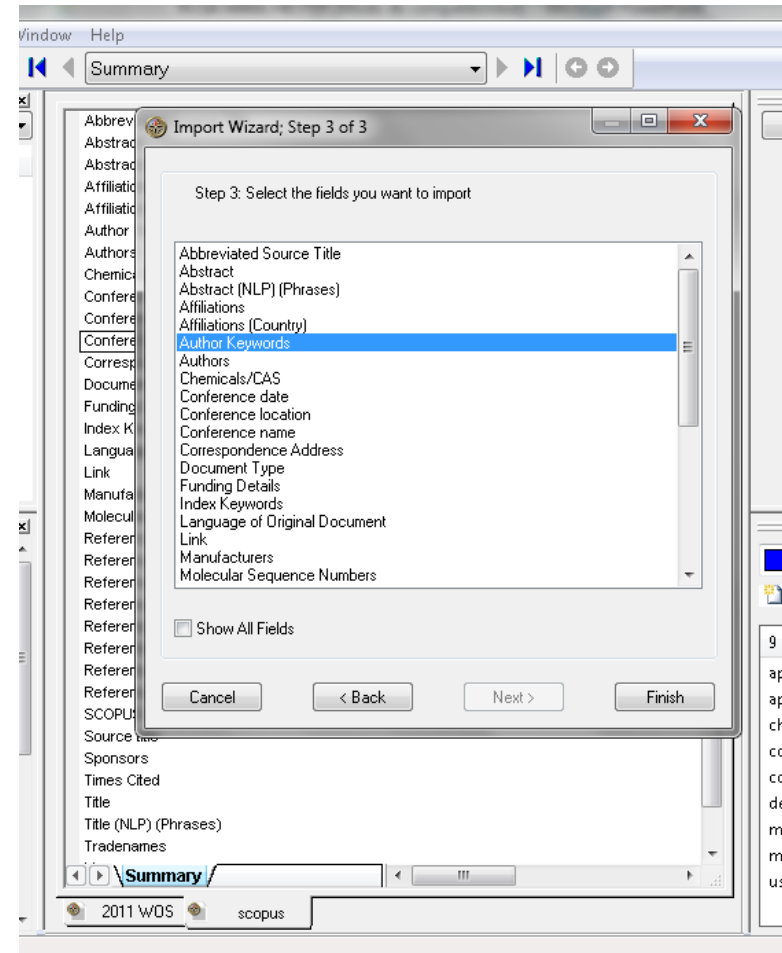
The analysis will use author keywords as cognitive units reflecting the research taking place in waste recycling field.

The merging process produced a VP file containing only article titles and author keyword field, in order to minimize file size and at the same time, keep fields (title) that could be further used to detect duplicities.

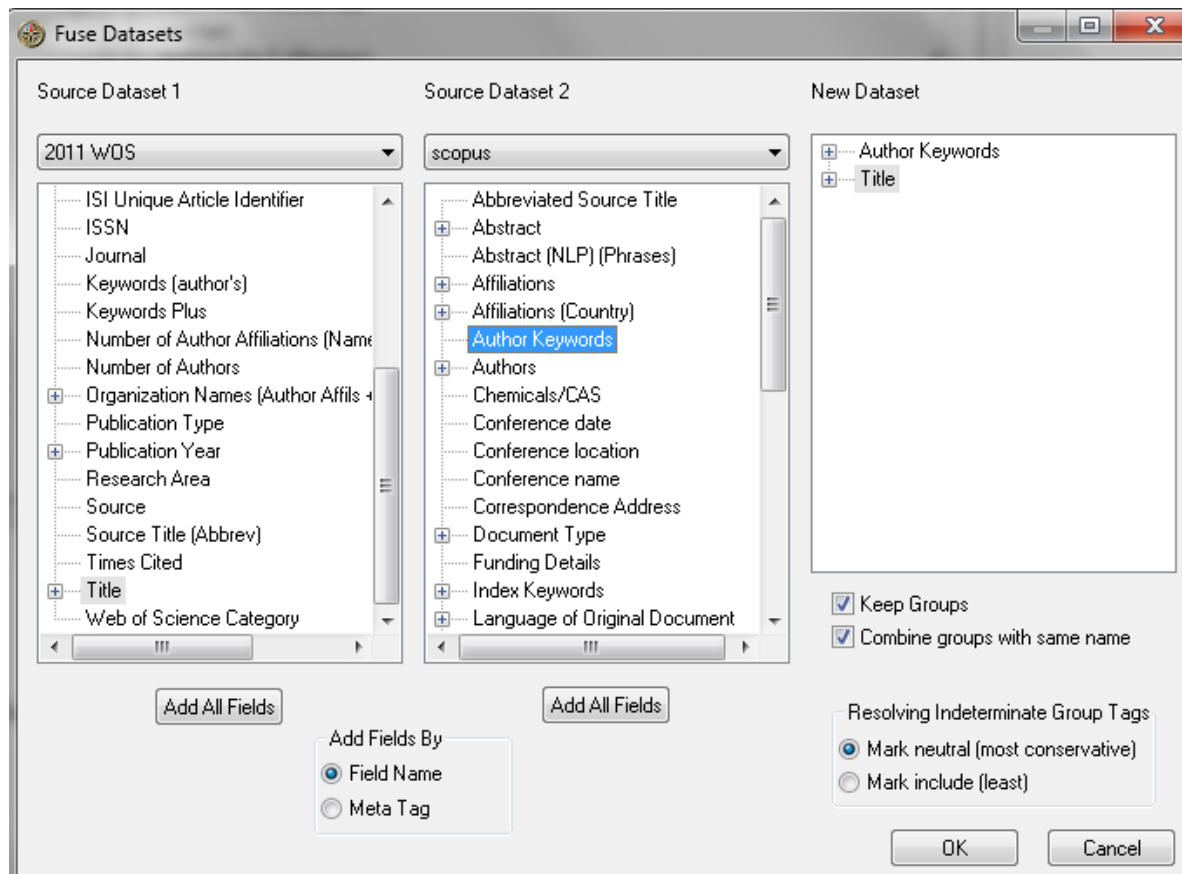
4th Step

Data import and merging.

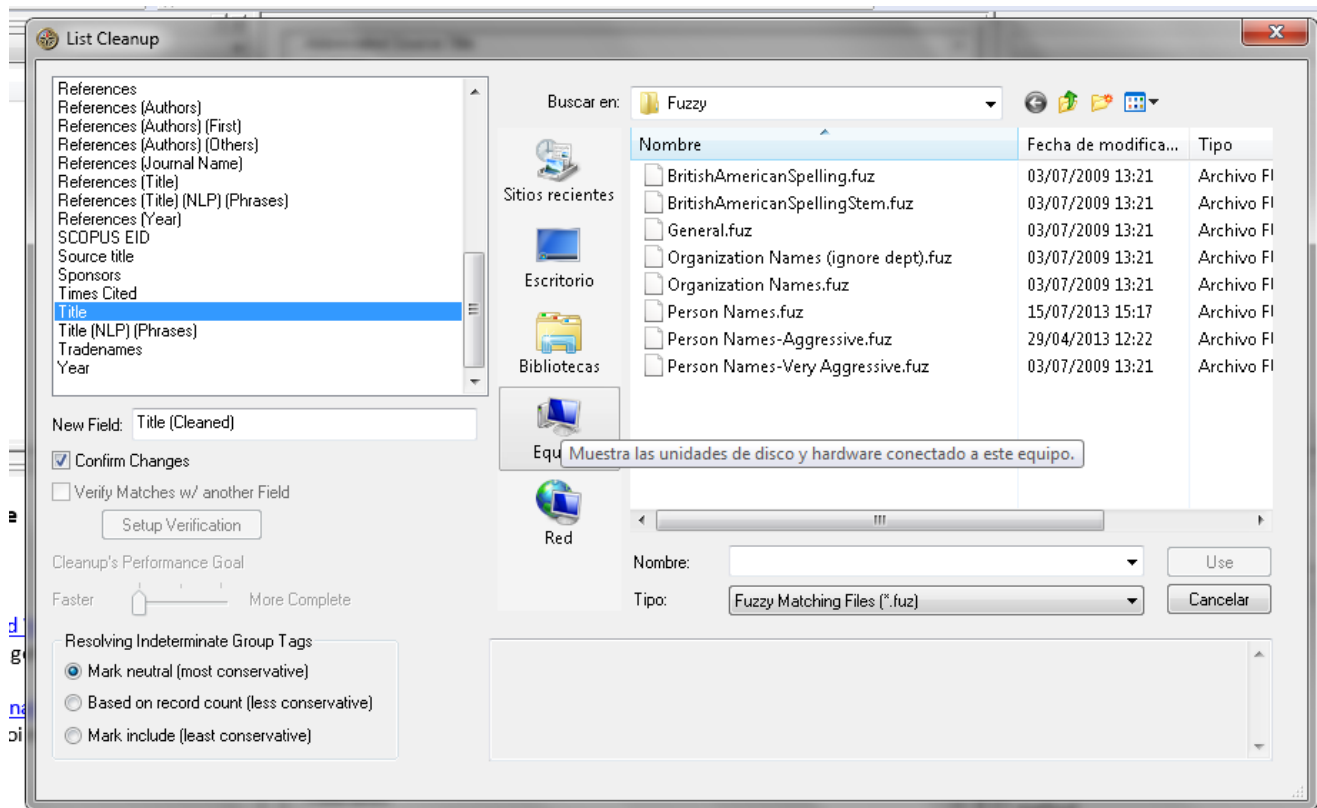
- Import wizard was used for importing data, using scopus(csv).conf and ISI-WOS.conf filters.



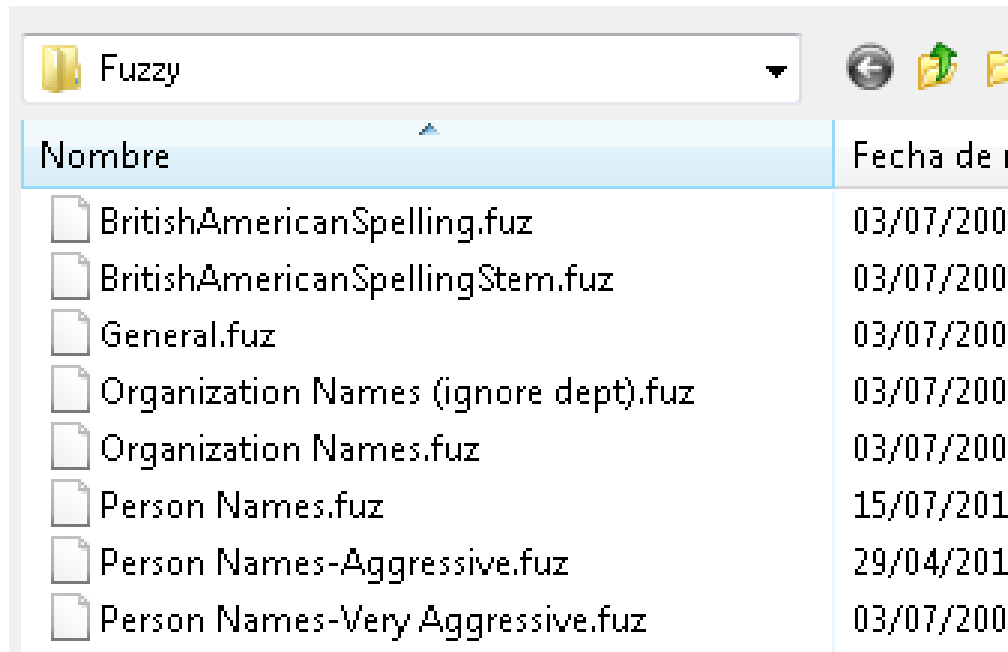
- Tools/Data Fusion command merges data from different VP files, allowing the user to choose the fields to merge.
- A good deal of duplicities are automatically detected by VP in merging process.



- Our main problem were the duplicities derived from the merging of the contents of several databases.
- «Title» field was cleaned by running several «list cleanup» commands

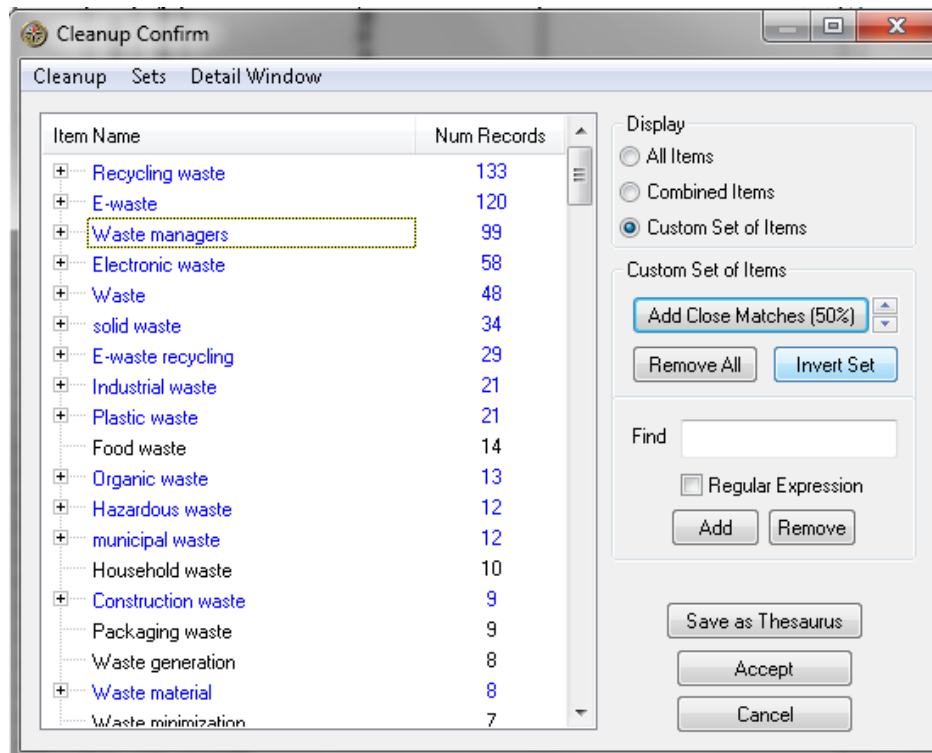


- There are many fuzzy matching files to detect duplicities, our approach was to start by running «General.fuz» (conservative), to later expand the cleaning with other fuzzy files that properly detected title variations due to greek symbols, dashes and other special characters. The cleaning must be conducted under close supervision of the analyst, since automatic cleaning is prone to errors.

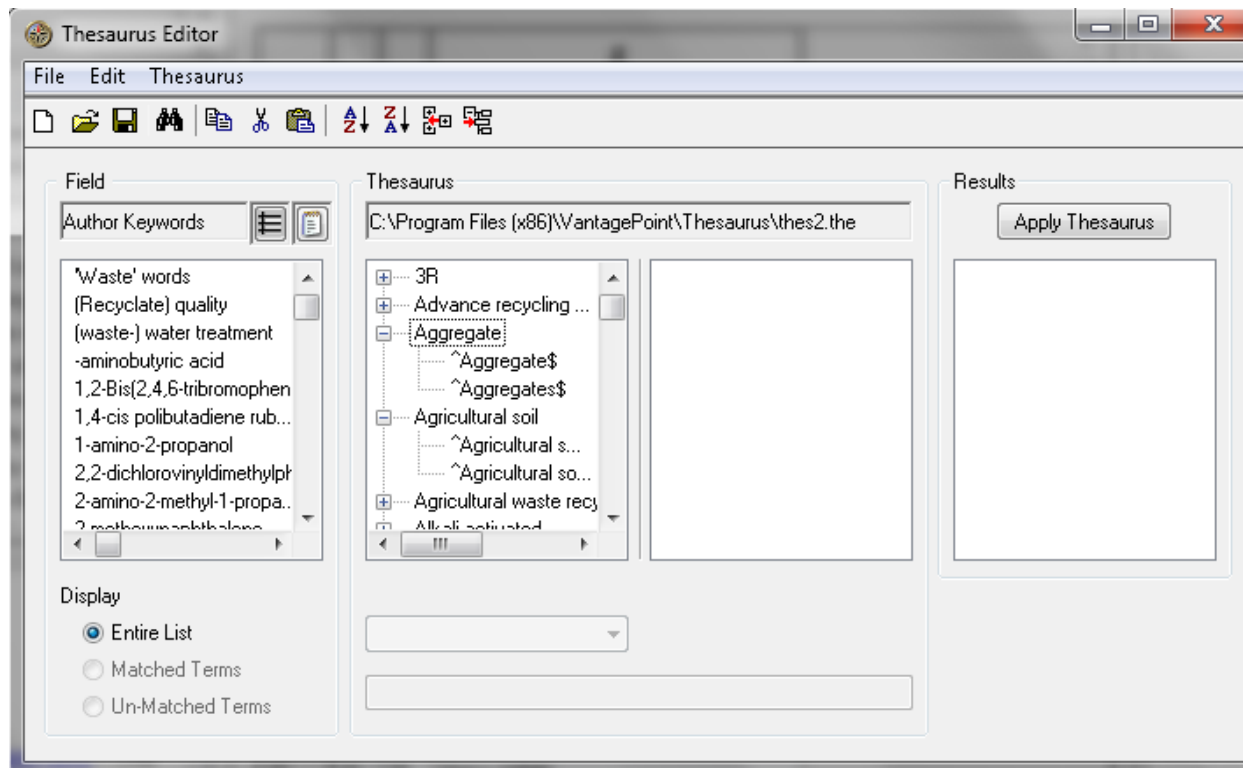


Nombre	Fecha de r
BritishAmericanSpelling.fuz	03/07/200
BritishAmericanSpellingStem.fuz	03/07/200
General.fuz	03/07/200
Organization Names (ignore dept).fuz	03/07/200
Organization Names.fuz	03/07/200
Person Names.fuz	15/07/201
Person Names-Aggressive.fuz	29/04/201
Person Names-Very Aggressive.fuz	03/07/200

- Having removed duplicities, then «author keyword» field was to be cleaned. In this case singular/plurar forms, synonyms and corrupted forms of the same term were to be grouped. «General.fuz» file and the option «Add Close Matches» were extensively used, by manipulating the similarity % in successive cleaning rounds.



- It is extremely important to build a thesaurus in this cleaning stage, since this thesaurus can be automatically run on the data corresponding to other databases or years.
- VP allows to build a thesaurus containing the operations made in each cleaning round. When the process is finished, all the thesauri can be merged, forming a complete thesaurus.



6th Step

Generating co-occurrence matrix.

- Once «author keyword» field is properly cleaned and duplicities removed, the co-occurrence matrix is easily created in VP, allowing the calculation of relationships between bibliometric items.

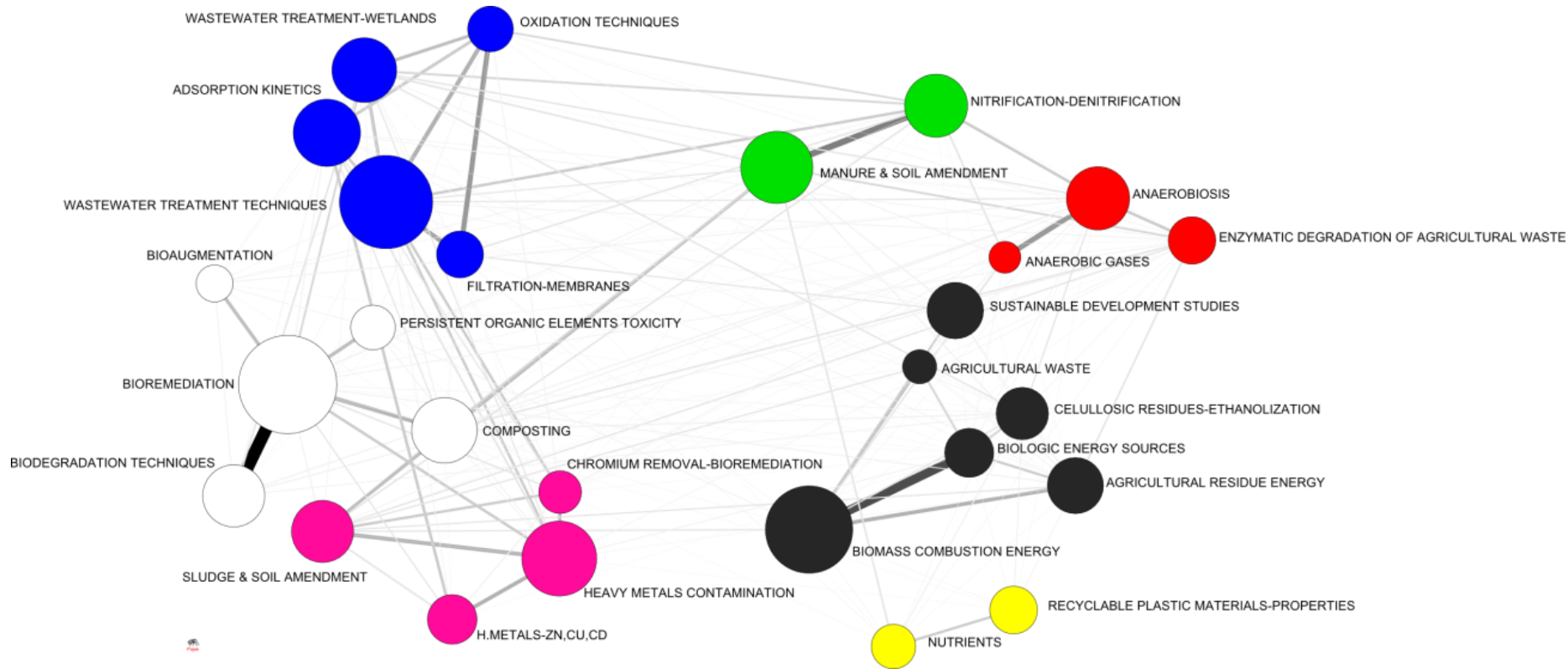
Reset	Keywords (author's) thes2810 AGRSv10	# Records	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
		▼ ▲																		
		Show Values >= 1 and <= 1083																		
		Cooccurrence # of Records																		
		▼ ▲																		
			WASTE-WATER TREATMENT	Biomass	Energy	HEAVY METALS	soil	MODELING	Biodegradation	ANAEROBIC DIGESTION	adsorption	Sewage Sludge	Kinetics	Bioremediation	COMPOST	MUNICIPAL SOLID WASTE	MANURE	biofuels	Biogas	
1	1083	WASTE-WATER TREATMENT	1083	13	26	45	19	41	21	23	44	28	23	6	4	5	1	11	13	8
2	1012	Biomass	13	1012	116	10	14	37	2	16	11	10	45	3	5	6	6	56	24	7
3	591	Energy	26	116	591	2	2	18	3	23	1	8	2		2	20	6	31	45	1
4	501	HEAVY METALS	45	10	2	501	51	8	1	6	39	52	9	25	16	21	7	1	4	2
5	501	soil	19	14	2	51	501	8	28	3	5	25	4	35	29	4	39	7	4	4
6	452	MODELING	41	37	18	8	8	452	7	18	9	5	27	9	10	15	8	12	10	7
7	425	Biodegradation	21	2	3	1	28	7	425	12	7	3	19	34	8	7	7	2	10	1
8	401	ANAEROBIC DIGESTION	23	16	23	6	3	18	12	401	3	30	13	1	11	17	34	3	87	
9	386	adsorption	44	11	1	39	5	9	7	3	386	10	53	3	2	1		1		7
10	371	Sewage Sludge	28	10	8	52	25	5	3	30	10	371	10		26	8	4	1	10	1
11	365	Kinetics	23	45	2	9	4	27	19	13	53	10	365	3	4	1	7	4	3	3
12	353	Bioremediation	6	3		25	35	9	34	1	3		3	353	1		1	2		
13	334	COMPOST	4	5	2	16	29	10	8	11	2	26	4	1	334	21	29	2	4	1
14	330	MUNICIPAL SOLID WASTE	5	6	20	21	4	15	7	17	1	8	1		21	330	4	1	10	5
15	324	MANURE	1	6	6	7	39	8	7	34		4	7	1	29	4	324	1	18	2



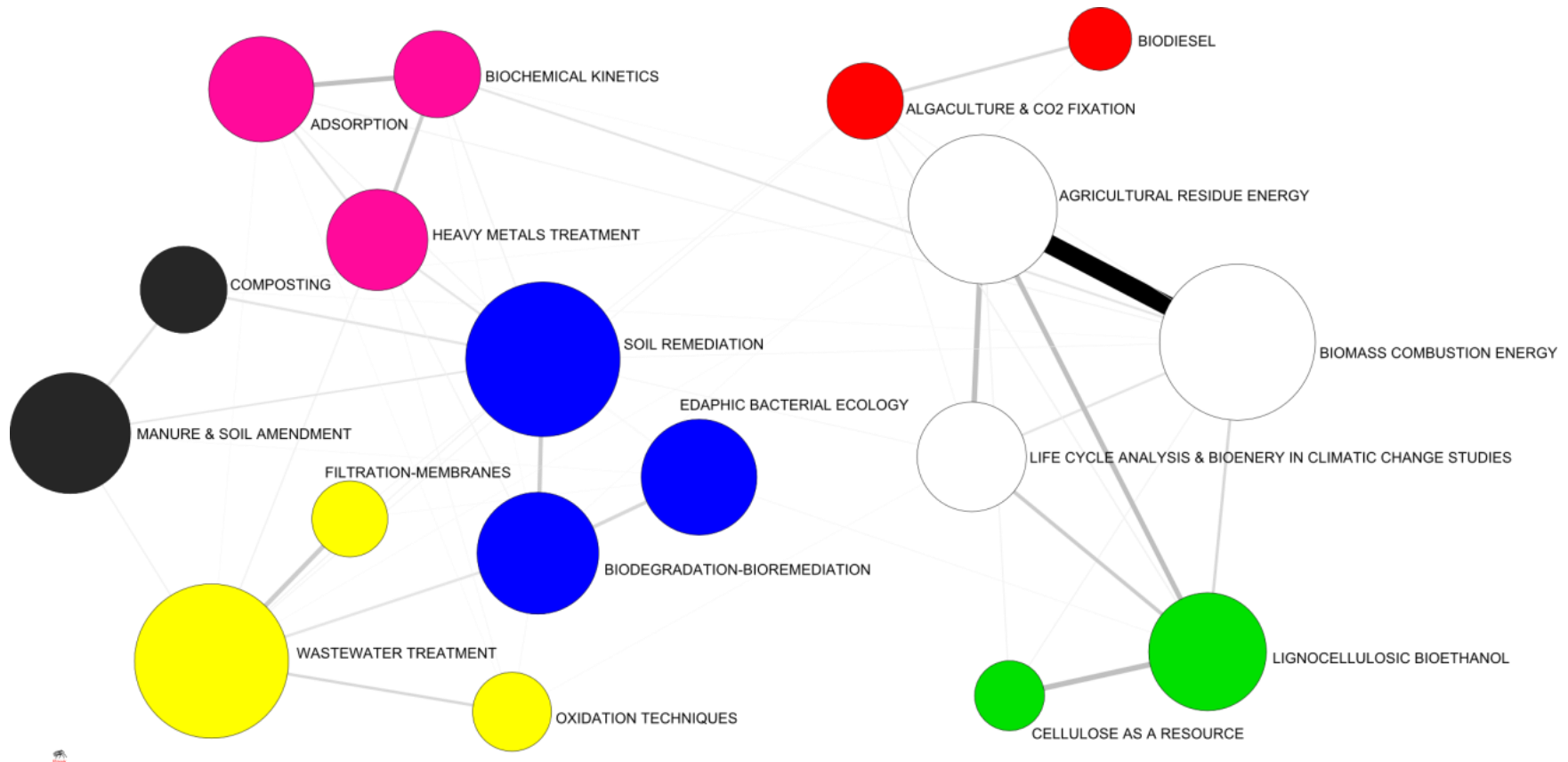
7th Step

Visualizations.

- **VP offers a wide variety of similarity calculation tools, as well as visualization tools, but the approach chosen in this study required the export of co-occurrence matrix for further processing to statistical software R.**
- **One of the weak points of this software lies in its problematic to export large matrices to a format that could be imported to R or other software. (slow 1000 x 1000)**
- **A similarity measure was calculated in R and keywords were clustered by hierarchical clustering. The clusters were further analyzed by network analysis using pajek.**



Map corresponding to main research areas in 2002. Each node corresponds to a keyword cluster, labelled by expert-supported analysis of the keywords. Links between nodes indicate similarity and node colours identify strongly connected clusters that form a wider research area.



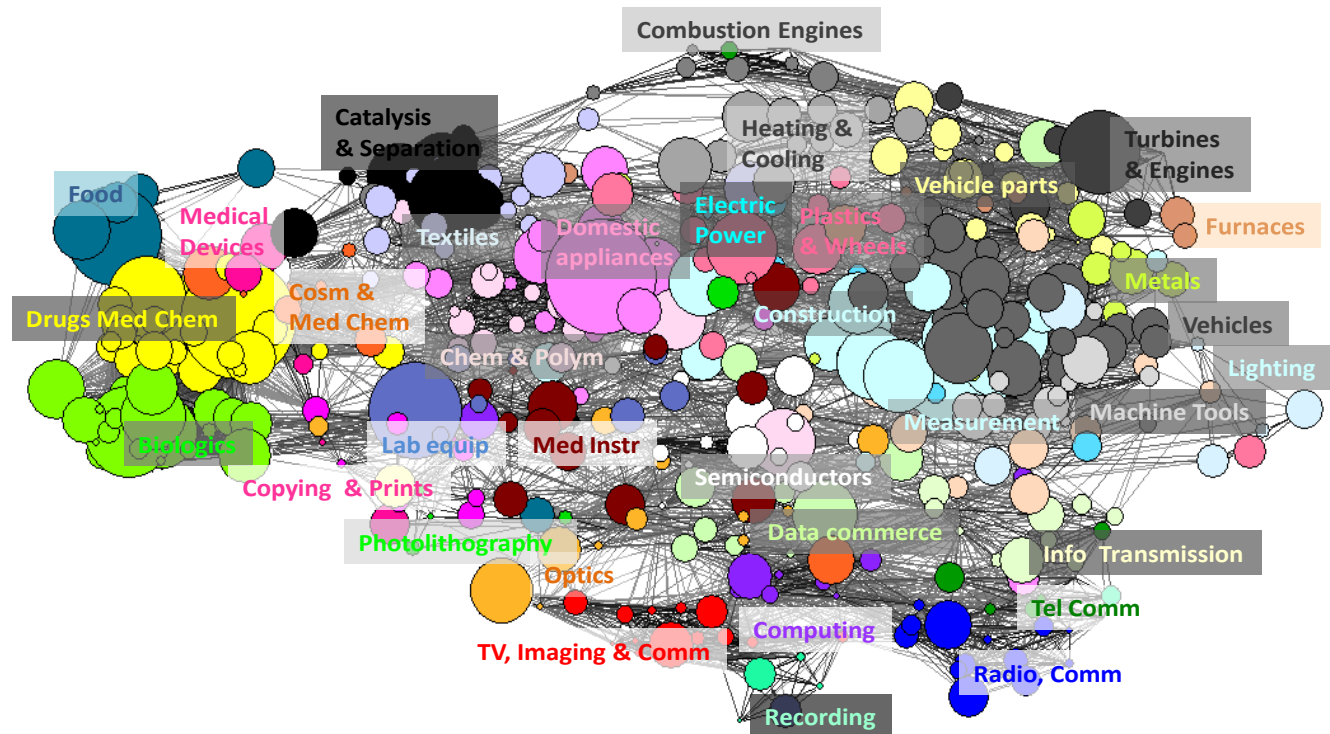
Map corresponding to main research areas in 2012. Each node corresponds to a keyword cluster, labelled by expert-supported analysis of the keywords. Links between nodes indicate similarity and node colours identify strongly connected clusters that form a wider research area.

This study uses the new global patent map developed by Luciano Kay et al. to reflect the patenting activity of Spain together with the activity of the Basque Country, a highly industrialized region in Spain.

The global patent map reflects the technology categories where a patent could be categorized according to the International Patent Classification (IPC) system, in addition to the degree of similarity among different IPCs, determined by using the citing-to-cited relationships as bonds between categories. An overlay method has been developed to compare both regions representing the most important technology fields and possible technology transfers.

The period of the study corresponds to Jan 2000 to Dec 2006, coinciding with the period of the global patent map.

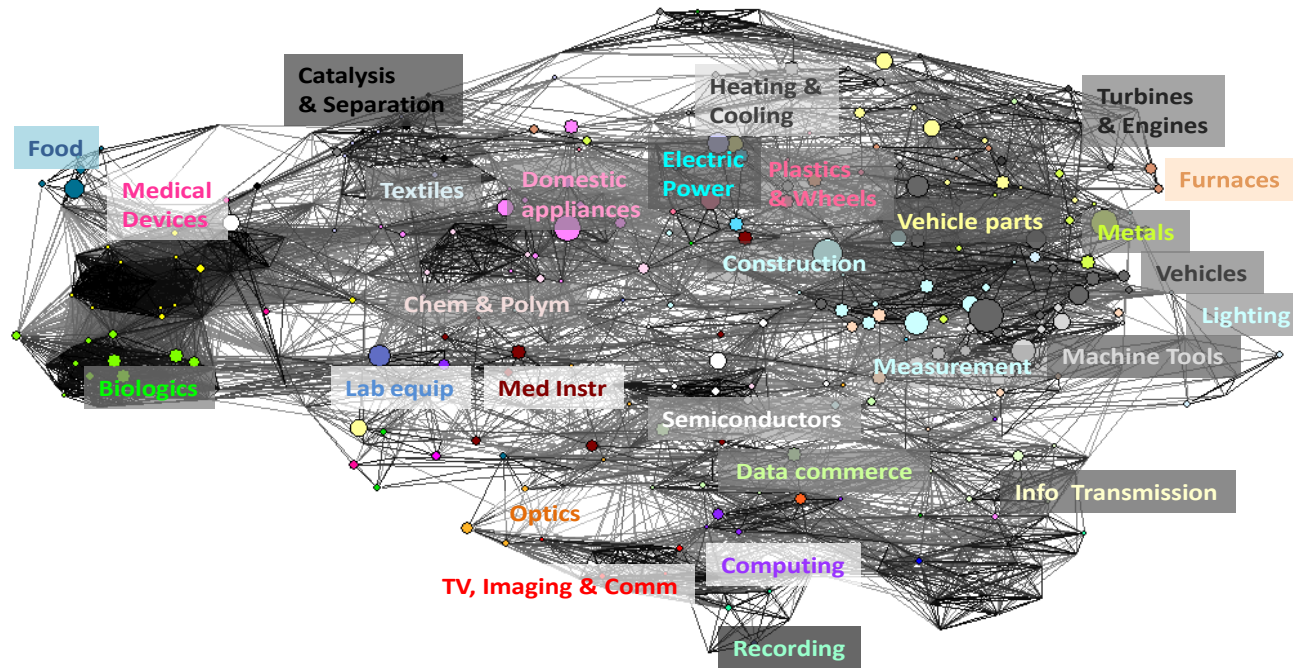
Spain Patent Overlay Map



Spain overlay has been made utilizing data corresponding to Spanish patent activity, collected from the PATSTAT database of European Patent Office (EPO) by using the nationality of applicants as selection criteria (13575).

Each node represents each of the 466 categories that simplify the IPC, and each colour represents each of the 35 technology areas in which they have been grouped. Sectors with higher inventive activity are: "Construction"; "Domestic appliances"; "Vehicles"; "Drugs, Med Chem" and "Biologics".

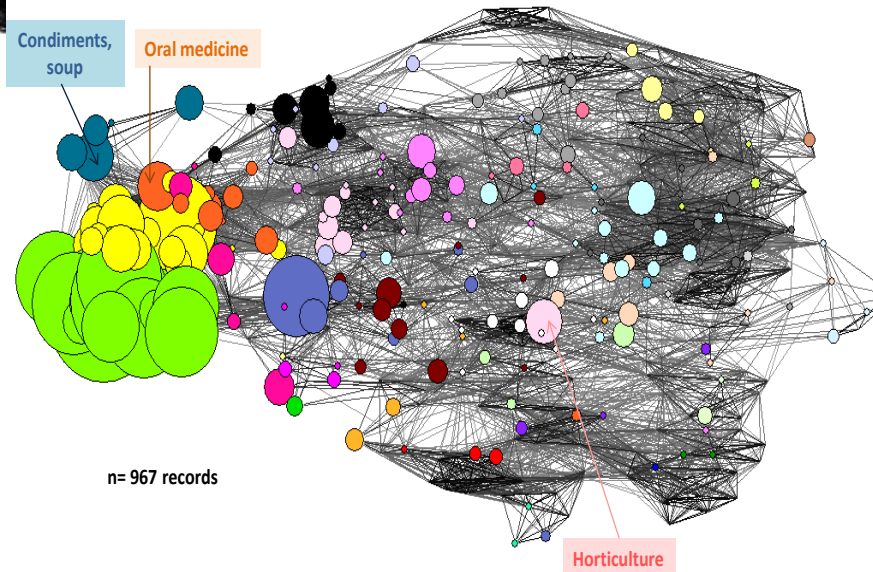
Basque Country Patent Overlay Map



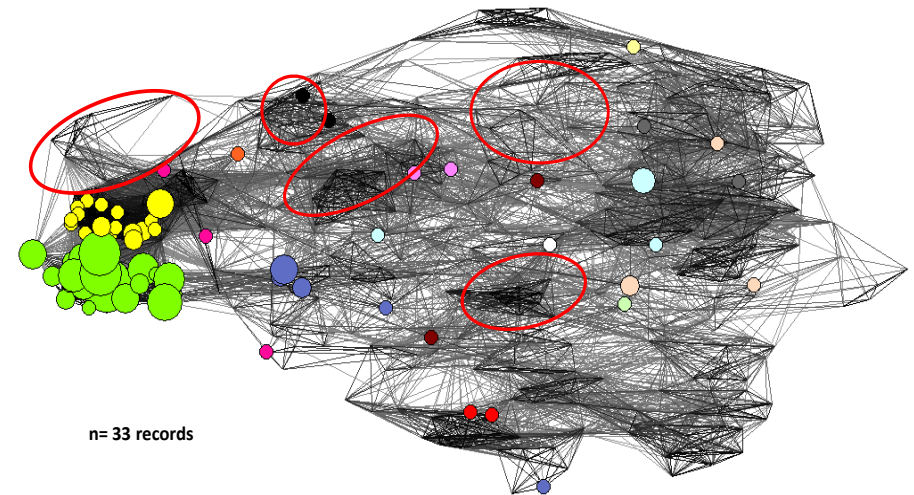
Basque Country Patent Overlay Map has required access to the INVENES database of Spanish Patent and Trademark Office (OEPM) in order to determine the region corresponding to each Spanish patent.(1038). Most important sectors coincide “Construction”, “Domestic appliances” and “Vehicles”; but “Vehicle parts” and “Machine Tools”, which are not very important in the case of Spain, also appear.

*Technological
knowledge flows in the
sector of “Biologics”*

Spanish cited Patent Overlay Map “Biologics”



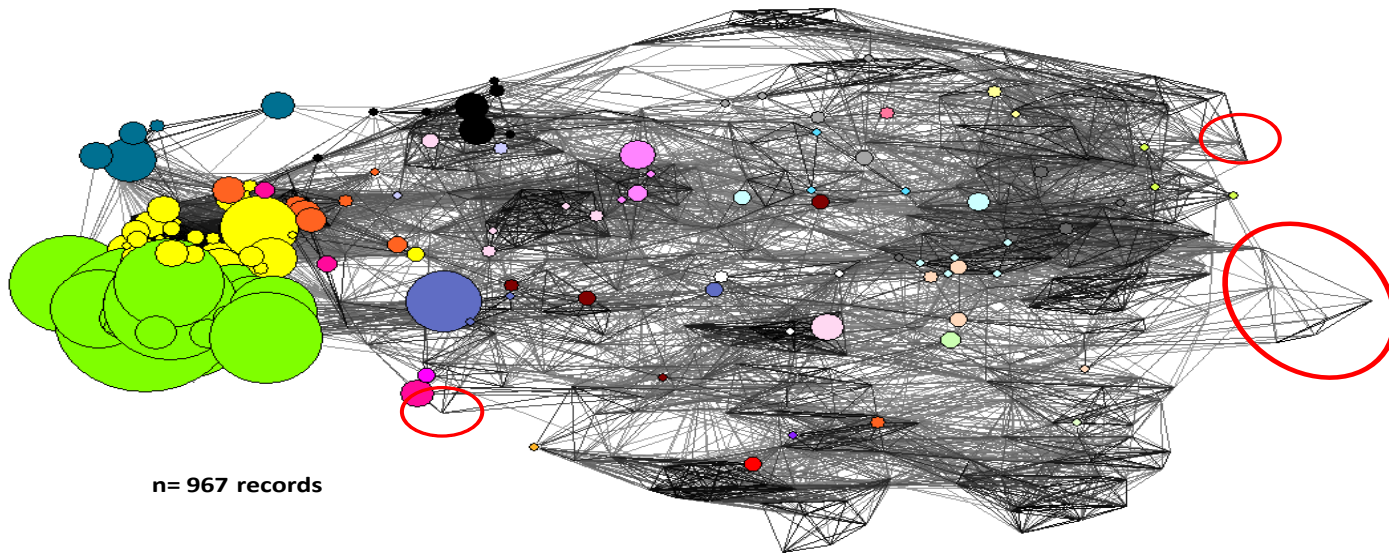
Basque Country cited Patent Overlay Map “Biologics”



If the patents related to the sector of “Biologics” are analysed only through their IPCs, and the IPCs that are cited are represented, the following overlays are obtained.

If they are compared, it can be observed how in the case of the Basque Country there are empty zones, which shows a shortage of technological flow among certain IPC categories in the sector of “Biologics” that are not met in the case of Spain: “Condiments, soup”; “Oral medicine” and “Horticulture”

.Spanish Patent Overlay Map “Biologics”



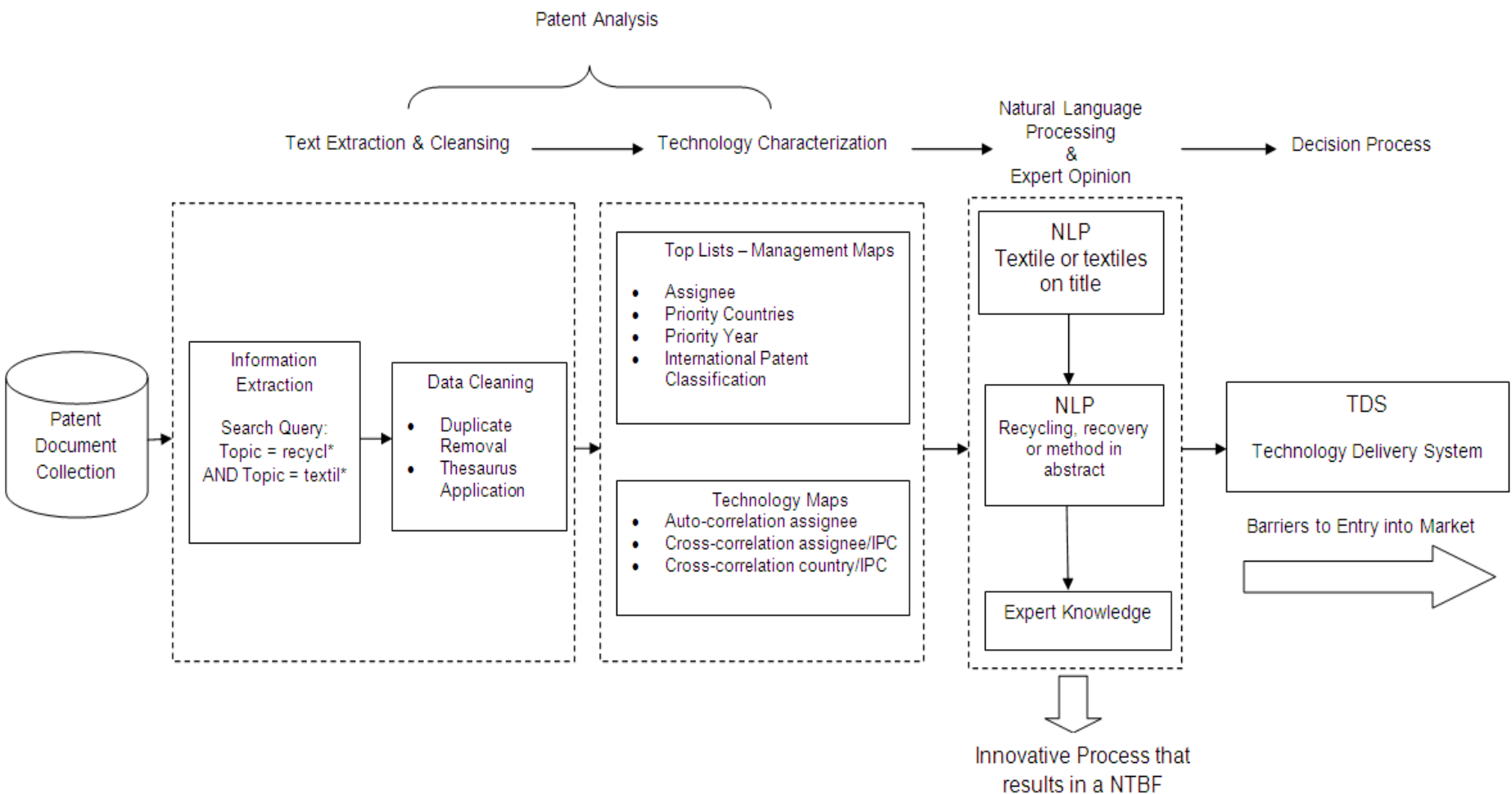
The patents related to the sector of “Biologics” are also related to other sectors
If this overlay is compared to that of the sectors to which the cited patents
belong to, it can be observed how certain sectors belong to the cited
patents but not to the citing patents: “Photolithography”; “Lighting”; “Furnace

- Firstly, a patent analysis in the field of textile waste recycling is performed, the main objective of which is to gain insight into technological trends.
- Secondly, once the technology landscape has been shaped, we proceed with the selection of the right patent which will be the grounds for the business proposal of a NTBF. The aim of this business proposal will be to set up a pre-treatment plant for post-consumer carpets generated in the Basque Country.
- Finally, the design of the Technology Delivery System that will allow us to identify, on the one hand, potential barriers when entering the market, and on the other hand, the major players along with the main leverage points that connect these emerging technology capabilities to market needs.



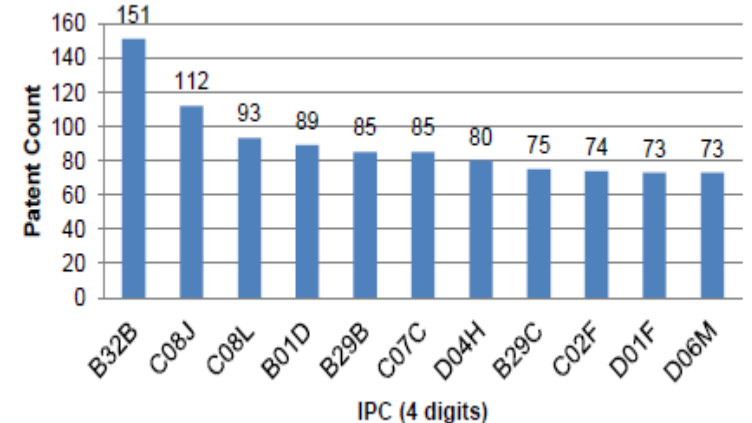
Patent Analysis study to Create New Technology Based Firms

In this case the patent information was retrieved from the Derwent Innovations Index database. The information retrieved is made up of 1156 patents found in the world textile waste recycling sector for the period 1965-2010.



Patent Analysis study to Create New Technology Based Firms

IPC s TOP 11



B32B Layered products, i.e. products built-up of strata of flat or non-flat, e.g. cellular or honeycomb, form
 C08J Working-up; general processes of compounding; after-treatment not covered by subclasses
 C08B, C08C, C08F, C08G or C08H C08L
 Compositions of macromolecular compounds
 B01D Separation
 B29B Preparation of pretreatment of the material to be shaped; making granules or performs; recovery of plastics or other constituents of waste material containing plastics
 C07C Acyclic or carbocyclic compounds
 D04H Making textile fabrics, e.g. from fibres or filamentary material; fabrics made by such processes or apparatus, e.g. felts, non-woven fabrics; cotton-wool; wadding

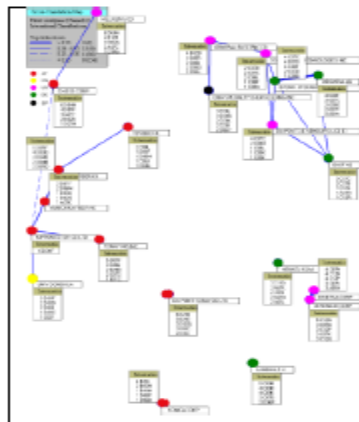


Figure 8a: cross correlation assignee/IPC 2000-

2009. 611 patents

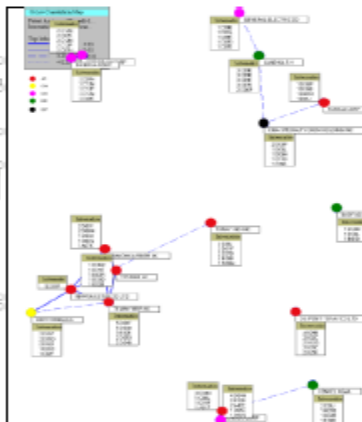


Figure 8b: cross correlation assignee/IPC 2000-2009

textile or textiles on title. 259 patents

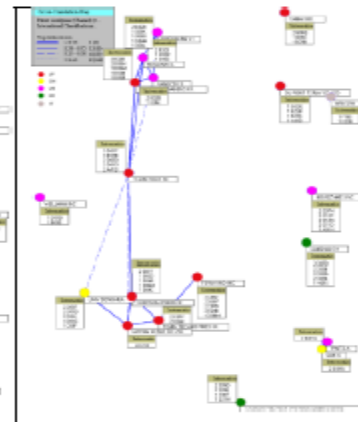


Figure 8c: 2005-2009 cross correlation assignee/IPC

textile or textiles on title. 147 patents

TOP 21	Number of Patents	Patent Assignee
1	30	TEIJIN FIBRE KK
2	30	TORAY IND INC
3	9	DU PONT TORAY CO LTD
4	8	DAIICHI AG
5	8	GENERAL ELECTRIC CO
6	7	HIKARI KOGA
7	7	MELLIKEN & CO
8	6	CHERO CORP
9	6	NIPPON KASEI CO LTD
10	5	CB&I SPECIALTY CHEM
11	5	NOVOLINE INC
12	5	DAIICHI AG
13	5	DAIICHI AG
14	5	LANXESS H
15	5	TOYOBO KK
16	5	UNIV DOKUHA
17	5	VERBODEN CORP
18	4	VERBODEN & SONS NV

TOP 17	Number of Patents	Patent Assignee
1	9	TEIJIN FIBRE KK
2	6	TORAY IND INC
3	5	DU PONT TORAY CO LTD
4	5	NIPPON KASEI CO LTD
5	5	YUNESKEA FIBRE KK
6	4	MELLIKEN & CO
7	4	UNITIKA LTD
8	4	UNIV DOKUHA
9	3	CHERO CORP
10	3	CB&I SPECIALTY CHEM
11	3	HOLDING INC
12	3	DAIICHI AG
13	3	DAIICHI AG
14	3	LANXESS H
15	3	MITSUBISHI KAYON CO LTD
16	3	TEIJIN TRICHO PROD KK
17	3	VERBODEN S

table 3. Figure 8b

TOP 18	Number of Patents	Patent Assignee
1	6	TEIJIN FIBRE KK
2	5	TORAY IND INC
3	4	NIPPON KASEI CO LTD
4	4	UNIV DOKUHA
5	4	YUNESKEA FIBRE KK
6	3	HIKARI KOGA
7	3	AKA SPA
8	3	DOKUHA V I
9	3	DU PONT TORAY CO LTD
10	3	LANXESS H
11	3	MELLIKEN & CO
12	3	NOVOLINE INC
13	3	DAIICHI AG
14	3	SUN X
15	3	TAMAKI KK
16	3	TEIJIN TRICHO PROD KK
17	3	VERBODEN S
18	3	WACKER POLYMER SYSTEMS

table 4. Figure 8c

Cross correlation assignee/IPC 2000-2009 textile or textiles on title. 259 patents

2005-2009 cross correlation assignee/IPC textile or textiles on title. 147 patents

Cross correlation assignee/IPC 2000-2009. 611 patents

Other projects



TFM research group

Industrial Organization and Management Engineering Department
University of the Basque Country (UPV/EHU)

(2012-2013) CLOUDROAD. Roadmap to Cloud Computing in the SME _

Programme: SAIOTEK 2012

Project Reference: SAI12/118

Funding Entity: Gobierno Vasco

Research Line: Cloud Computing: Business Perspective (UNESCO code: 530600)



(2012-2013) Forward in the visualisation of the connection between Science and Technology

Project Reference: EHU12/19

Funding Entity: UPV/EHU

Research Line: Text Mining Technology /Techmining (UNESCO code: 530600)



(2011-2013) Smart Platform of Business Management based in Cloud Computing y la WEB 2.0 _

Programme: INNPACTO 2011

Project Reference: IPT-2011-1805-430000

Funding Entity: Ministerio de Ciencia e Innovación

Main Researcher: Instituto de Innovación Empresarial S. A.

Research Line: Cloud Computing: Business Perspective (UNESCO code: 530600)



Thank you very much



CAMPUS OF
INTERNATIONAL
EXCELLENCE

Río-Belver, R.M.
Rosamaria.rio@ehu.es

[https://sites.google.com/
site/tfmresearch/tfm](https://sites.google.com/site/tfmresearch/tfm)