



EDP in an emerging field - text mining for the identification of academic experts for Industry 4.0 in Toscana (Italy)

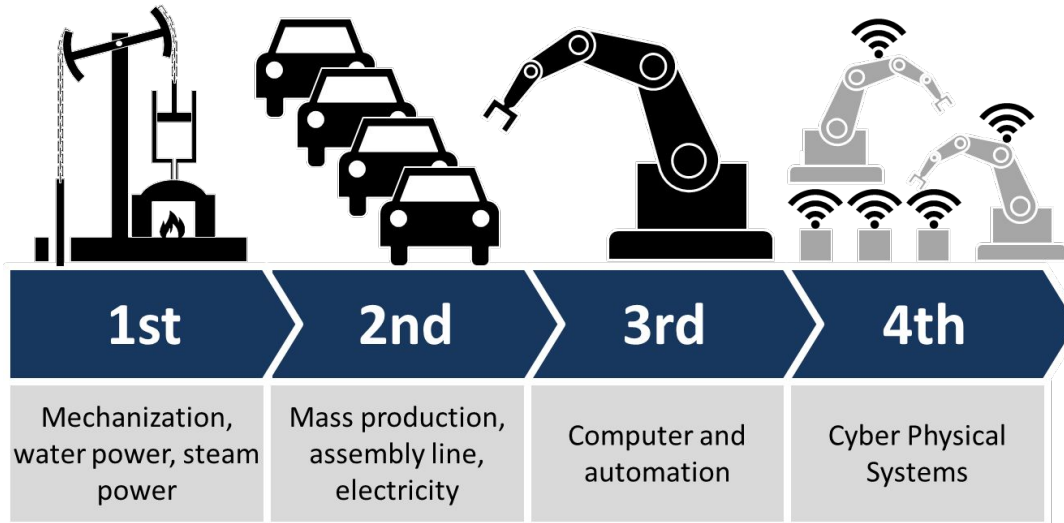
Francesco A. Massucci¹, Albino Caporale², Manuela Bigi², Caterina Brizzi², Silvia Burzagli², Enric Fuster¹, Alessandro Mosca¹, Bernardo Rondelli¹

¹SIRIS Academic SL, Av. Francesc Cambó 17, 08003, Barcelona, Spain

²Regione Toscana, Direzione Attività Produttive, Via Luca Giordano 13, 50132, Firenze, Italy

Industry 4.0 (I4.0), a quick introduction

Industry 4.0 (I4.0) is the name associated to the introduction of digital technologies and data exchange in manufacturing processes.



The paradigm shift in manufacturing technologies is comparable to what achieved with the introduction of:

- steam power
(1st industrial revolution)
- mass production
(2nd industrial revolution)
- Automation
(3rd industrial revolution)

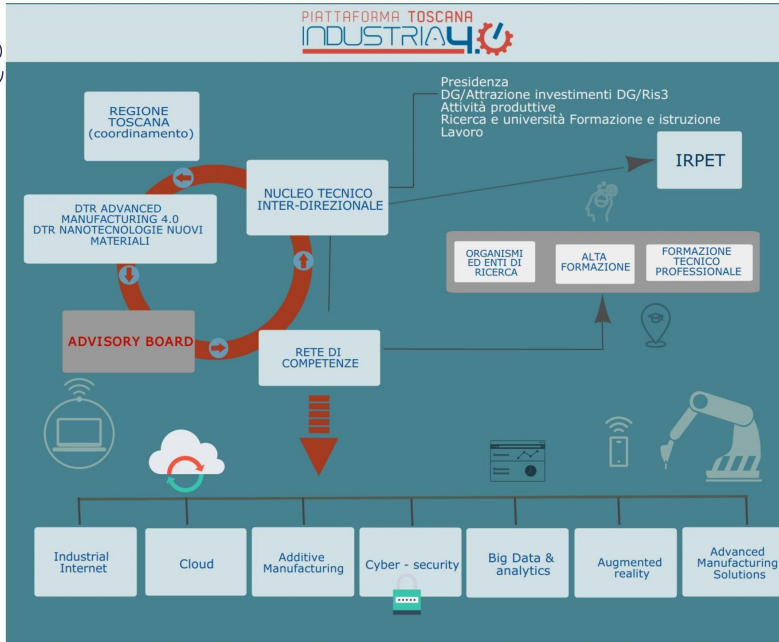
The national context:

The Italian *Ministry of Economic Development's* strategies on I4.0

Investments, fiscal deductions and incentives to

- Enhance private R&D investments
- Ease the creation of network infrastructures, data protection, interoperability standards
- Provide platforms for building and matching skills (R2B, B2B)
- Raise awareness and foster governance at public/private interface

The regional response: The I4.0 platform in Tuscany



- Analyse I4.0 needs of private actors;
- Identify local I4.0 skills (public/private)
 - *AUGMENTED REALITY*
 - *BIG DATA & ANALYTICS*
 - *CYBER SECURITY*
 - ...
- Integrate skills in regional lifelong learning programs;
- Match I4.0 demand/offer;
- Help in designing policies and instruments to facilitate the 4.0 transition.

In line with the **EC Smart Specialisation Platform** goal of Increasing Industrial Competitiveness in EU regions by raising awareness of Companies (SMEs) for Industry 4.0



I4.0: Research to business (R2B)

The needs of the Region:

1. Implementation of a directory allowing for:
 - The profiling of researchers by I4.0 technologies
 - The showcasing of research activities by researchers
 - The screening & selection by the Region

Of researchers and companies

2. Organisation of matching events
3. Practical matching of demand & offer

How to initially populate the directory database with researchers identified via their I4.0 skills?



Regione Toscana



GIOVANI SI

**TOSCANA
TECHNO
LOGICA**





EDP: getting in touch with stakeholders

Tuscan Higher Education & Research ecosystem, the basin for I.40 skills

15 Universities and research centres

277 Technology transfer structures

7000 Researchers

+ Regional cluster on Advanced Manufacturing

The problem at hand is:

How can one **identify & classify relevant actors for the EDP within this basin?**



EDP: getting in touch with stakeholders

EDP based on 'classical' participatory methodologies

Regione Toscana circulated a survey to Universities, research centres

But it's difficult to get hold of actors in the sphere of **Higher Education & Research:**

- Complex and redundant structures
 - Hard to get access to the relevant actors
- Technical jargon
 - Misalignment between offer/demand vocabularies
- Transversality of technique
 - Relevant expertise may be in unexpected places



EDP: Emergent thematic niches

I 4.0 is a **new topic**, formed by several pre-existing disciplines and technological domains: difficult to define and certainly **not in existing taxonomies**

Scopus

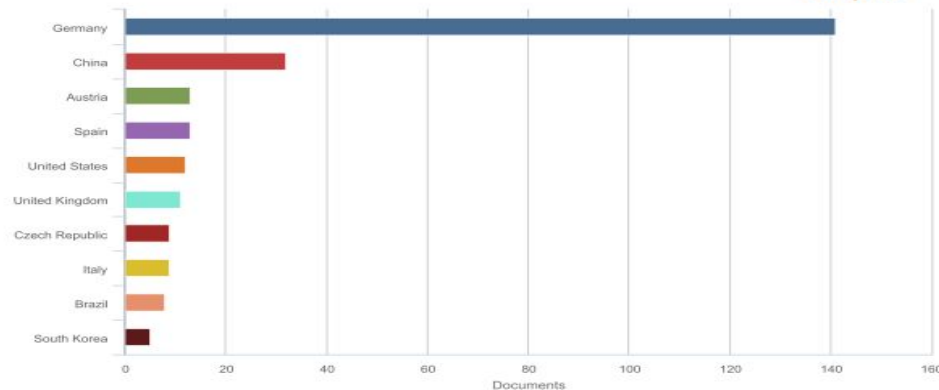
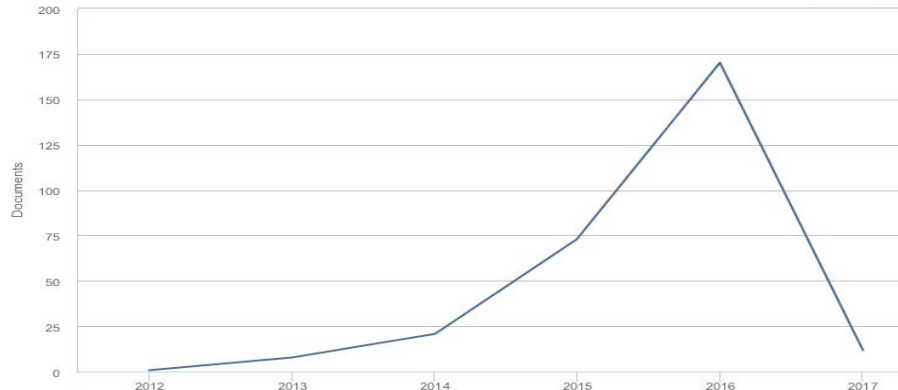
Search

Analyze search results

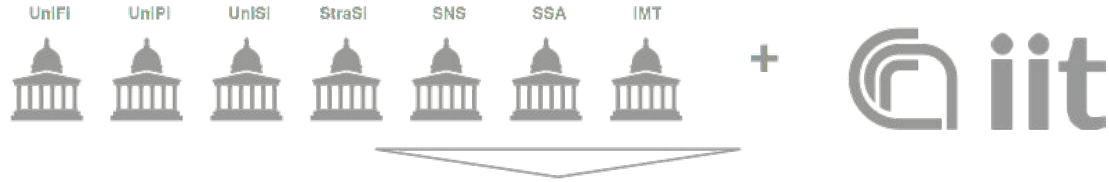
Analyze search results

KEY ("Industry 4.0") [Back to your search results](#)

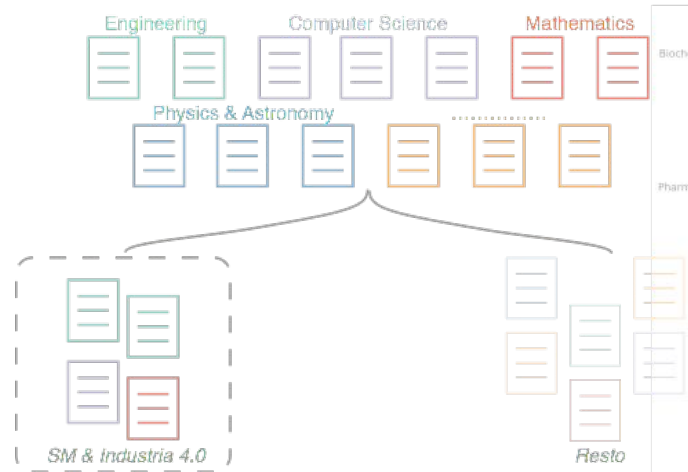
285 document results Choose date range to analyze: 2012 to 2017



EDP: Emergent thematic niches



2010-2015: about 57,000 publications
in total, in all disciplines





EDP: getting in touch with stakeholders

Big data, new technologies (machine learning) may help:

The tuscan HE&R ecosystem produces ~ **10,000 publication per year**
text mining to identify relevant documents and respective authors/skills

‘Classical’ methods are always fundamental, human component is indispensable
Definition of the semantic perimeter and of the competences, results
assessment



Machine Learning as a tool in EDP

By exploiting a *seed corpus* on I4.0 and an automatic classifier, one can extract regional publications on I4.0 themes and to identify their respective authors





Seed Corpus selection

Definition of the semantic perimeter and related I4.0 skills

- Scientific publications from EC *Factories of the Future* programme
- Scientific publications indexed by keywords:
 - “*Smart Manufacturing*”
 - “*Advanced Manufacturing*”
 - “*Industry 4.0*”

	N. documents
Factories of the future	484
K-word ‘Smart Manufacturing’	82
K-word ‘Advanced Manufacturing’	551
K-word ‘Industry 4.0’	226
TOTAL	1343

These texts fall outside the Tuscan production!



EDP: Text mining results

About

- 11,000 relevant publications identified
- 2000 researchers identified

Institution	#Researchers	#Pubs
Università di Pisa	735	4536
Università di Firenze	611	3344
Scuola Superiore Sant'Anna	201	1190
Università di Siena	162	1200
CNR ISTI	110	663
Scuola Normale Superiore	76	669
LENS	57	173
CNR IFAC	50	231
CNR INO	47	221
IMT Lucca	41	314
CNR IFC	24	...
EGO	21	...
CNR ICCOM	14	...

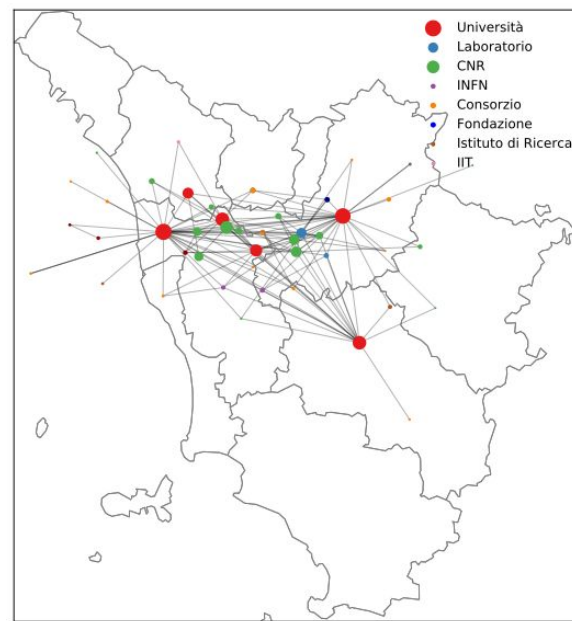
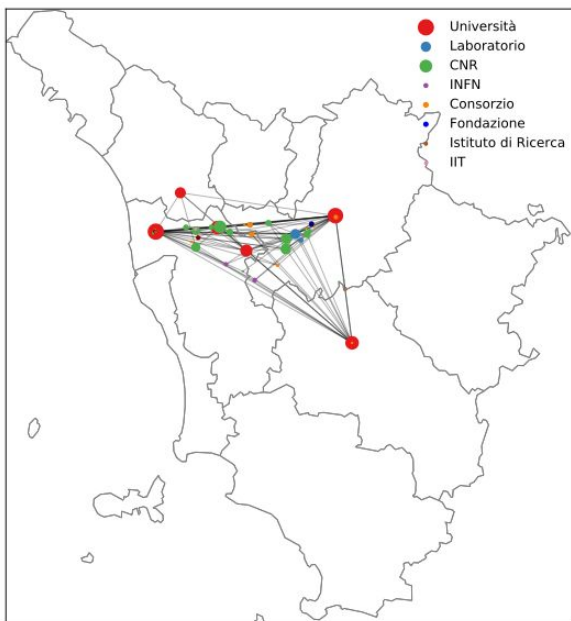
Institution	#Researchers
CNR IIT	11
CNR Nano	7
CNR ISC	5
Fondazione G. Monasterio	4
INFN	4
CNR ILC	4
CNR IVALSA	3
Osservatorio Arcetri	3
CNR IPCF	2
QSTAR	1
CNR IBIMET	1
INSTM	1
CNIT	1



EDP: Text mining results

About

- 11,000 relevant publications identified
- 2000 researchers identified



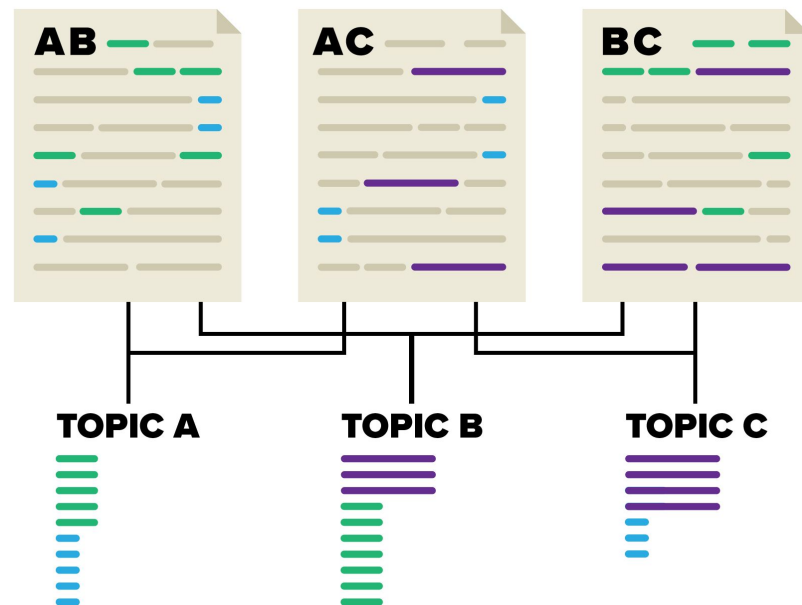


EDP: technological niches

Topic Modelling (TM) to identify potential technological domains

TM is a machine learning technique that returns a collection of topics that best reproduce the content of a given corpus.

Topic are a list of keywords that co-appear frequently in different documents





The extracted topics

#	Topic Label	Main words in topic	Importance
1	Big data and ICT	data, image, algorithm, analysis	15.7%
2	ICT for manufacturing	technology, process, product, advanced	10.8%
3	Design and testing of intelligent systems	performance, test, vehicle, safety	10.6%
4	Distributed computing	service, data, cloud, platform	8.3%
5	Theoretical computer science	function, problem, system, set	8.2%
6	Green energy production and storage	frequency, surface, power, device	7.1%
7	Algorithms	algorithm, solution, optimization, method	7.1%
8	Big data management	user, information, web, system	6.1%
9	Business models	market, analysis, research, italian	6.1%
10	Advanced processes for manufacturing	experimental, test, pressure, effect	5.9%
11	Multi-functional materials	magnetic, graphene, field, interaction	5.6%
12	Internet of Things	wireless, sensor, node, traffic	5.1%
13	Additive manufacturing	material, film, chemical, layer	4.3%
14	Natural language processing	analysis, language, logic, fuzzy	4.2%
15	Robotics	control, robotic, force, hand	3.8%
16	Sustainable Manufacturing	water, environmental, plant, production	2.7%
17	Bio nanomaterials	tissue, scaffold, protein, nanoparticles	2.6%
18	Advanced human-machine interaction	instrument, detector, telescope, energy	2.5%
19	ICT in medicine	clinical, medical, treatment, disease	2.3%



Results (I)

1. 2,200 researchers identified (including PhD students/untentured staff), each with a set of associated I4.0 skills

The Region performed quantitative and qualitative checks, to ensure

i. Researchers were still in those institutions,

ii. inferred I4.0 skills were correct.

About 1,000 professors confirmed!



2. About 200 regional companies were identified as publishing in collaboration



Results (II)

1. Based on these results, the region built the directory with specific areas dedicated to the showcasing of Labs, departments and single researchers
2. The directory is now used as a basis for the R2B events in the EDP contexts





Future work

1. Application of Artificial Intelligence on the dataset to
 - a. have a semi automated profiling
 - b. automatically I4.0 demand and offer
 - c. facilitate the align interaction between companies and R&D sector
2. Vertical analysis on public/private interface
3. A wider analysis including publications from companies
4. Benchmarking: comparing with regional competitors EU-wide
5. Including patents and/or research projects