INTELCOMP DATA TOOLS

A "teaser" presentation of how IntelComp will work in practice

Jeronimo Arenas Garcia Universidad Carlos III de Madrid



IntelComp for Policy Making

- A platform for evidence-based policy making
 - Using **open data** related to STI in the different domains
 - Ability to **analyze** massive amounts of data from different sources
 - Exploiting state-of-the-art Artificial Intelligence Models: deep learning, NLP
 - Expert validated models and tools for creating your own models
 - Offering **flexible interactive visualizations**, co-created with users
 - Creating Al-enriched dashboards and publishing them (Participation Portal)
- A workbench of tools for project evaluation
 - Search tools based semantic similarity
 - Evaluator assignment

Further advancing technologies already tested in two previous experiences: Corpus Viewer & Data4Impact



IntelComp Platform



Citizens / Academia

IntelComp front-end



IntelComp Platform: Model Trainer

- Definition of target datasets
- Selection of documents by target domain
- Training of models
- Tools for Curation & Annotation of Models
- Application of available models
- Configuration of predefined views

My Data Space Management





Living labs will deliver open models for the three target domains: AI, Cancer, Climate Change



IntelComp Platform: STI Viewer

Visualization tools co-created with the living labs Agile methodology (user in the loop)

- Topic and graph enriched BI dashboards
- Time and geographic STI analysis
- Comparison of different corpus (e.g., EU vs
- Profiling of STI agents: researchers, organizations
- Analysis of projects by instrument, management unit, etc

7.91% (0.696)

5.69% (0.726)

7.92% (0.712)

8.28% (0.694)

5.12% (0.736) 4.93% (0.696)

4.17% (0.725)4.87% (0.712)

6.27% (0.703) 6.28% (0.696)

7.58% (0.686)

6.79% (0.69)

6.75% (0.682)

• Exportation of results





AI-models vs Taxonomical Analysis

- Heterogeneous classifications, usually not compatible between distinct corpus (IPC-AJSC-Cordis classification-NSF/NIH class. ...).
- Problems with classification time evolution: speed of adaptation, backward compatibility (ex. IPC versions).
- Not suitable grain of definition on information (ex. HMI IPC codes).
- Long delay in the availability of data from statistics and surveys (sometimes 1-2 years)
- Poor document representation; binary multicomponent vector of classification labels (with few human defined components).
- Disjunctive classifications for business intelligence or statistical use; hybrid projects



A use case in the AI domain: Selection of Data Sources

- Identification of AI-related documents in the available data sources
- Aligned with other initiatives (SCOPUS, OECD)
- Taxonomies do not suffice for this task

Topic Modeling

Document

preselection

Expert knowledge



Lintelcomp

8

Policy Qs: Agenda Setting, Intelligence gathering, problem identification

Specific Policy Questions can be addressed by matching available data sets and AI-services

| Area | Po | icy Question | Data Sets | Int | elComp Services |
|-----------------------------|----|---|---|-----|--|
| Entrepreneurial Activity | • | Are companies adapting to technological transformation trends in their respective sectors? | Companies Websites Scientific papers | • | Detect technological transformation trends in a specific sector (i.e., AI techniques or AI applications) |
| | • | How do they compare with major (international) competitors? | Patents | • | Temporal evolution of topics Detection of lead-lags in |
| | • | Which companies emerge with specific disruptive technologies in the country/macroregion/region/city? | | • | Comparison of technologies across companies |
| | | | | • | Comparison of topic distribution in corpus by country |
| Knowledge Creation | • | Which scientific fields demonstrate the highest growth in terms of | Scientific papers | • | Distribution of topics in corpus |
| | | publications/citations globally? Distinction to be made between basic and applied research (distinction between interdisciplinary | Grants | • | Temporal evolution of topics (detection of "new" topics) |
| | | publications, basic research and applied research) (using journal classification?/calls for proposals) | (National,EU, International) | • | Impact analysis based on documents and authors (citations, topics) |
| | • | Which are the emerging interdisciplinary fields globally? | Patents | | |
| | • | Which are the research teams in the country undertaking research in these fields? | | | |
| ance | • | To which global, EU societal challenges (i.e. living lab specific) are research groups contributing to? | Scientific papers | • | Specific EU societal challenge identification in corpus (i.e. public grants) |
| Guid | | | Grants | • | Automatic classification of public funding proposals and grants by EU societal challenge based on a training set / challenge description |
| Market | • | What is the role of public procurement for these technologies | EC Public tenders | • | Identification of AI techniques and AI applications in corpus |
| | | (theoretically/practically)? | National Public tenders | | |
| Resources mobilization | • | What are the national/regional financial resources available in the | EU grants | • | Presence of AI techniques and AI applications in EU grants, national grants |
| | | country? Are they used to leverage EU funding through synergies? | National grants | • | Comparison of public funding in corpus (i.e. EU grants, national grants) by topic in a |
| | • | Is there a gap between supply and demand? | Job postings | | specific sector |
| | | | | • | Comparison of topic distribution in corpus (job supply, job demand). |

AI topics in Spain State Research Plan. Time Analysis. Region Analysis

0.04

0.06

0.08

0.02

Platforms, Services and Solutions (AI Functional Applications) Innovation and Technology (AI Framework and Strategy) Manufacturing (AI Field Applications) Mathematics and Statistical Modeling (Al Techniques) Modeling and Numeric Simulation (AI Techniques) Robotics (AI Functional Applications) Algorithms and Machine Learning (Al Techniques) Health: Healthcare (AI Field Applications) Software Engineering and Architectures (Al Functional... Decision Support: Algorithms and Models (AI Techniques and... Wireless Communicatons (AI Functional Applications) Transport (AI Field Application) Energy (AI Field Applications) Computer Vision and Imaging (AI Functional Fields) Signal Processing & Communications (AI Field Applications) Semantics and Ontology (AI Functional Fields) HW Architectures (Al Functional Applications) Health: Psychology and Cognition (AI Field Applications) Education (AI Field Applications) Infrastructure and Platforms (AI Functional Applications) Language Technologies (AI Functional Applications) Agriculture & Water Management (AI Field Applications) Health: Rehabilitation and Sports (AI Field Applications) Security & Biometrics (AI Field Applications) Astrophysics (AI Field Applications)



Which scientific fields demonstrate the highest growth? Distinctions between Basic and Applied Research
Which are the emerging interdisciplinary fields globally?
Which are the research teams in the country undertaking research in these fields?



elcon

10

- Platforms, Services and Solutions (Al Functional Applications)
- Innovation and Technology (AI Framework and Strategy)
- Manufacturing (AI Field Applications)
- Mathematics and Statistical Modeling (AI Techniques)

0,12

- Modeling and Numeric Simulation (AI Techniques)
- Robotics (AI Functional Applications)
- Algorithms and Machine Learning (Al Techniques)
- Health: Healthcare (AI Field Applications)

AI topics in FP7 and H2020 projects. Country distribution



Which are the countries to work with on the priority challenges?

Researcher / Research group Analysis & Profiling

- Based on project proposals, publications and patents
- Profiling based on citations, acknowledgments and topics
- Visualization connected to topics



Agente: CID SUEIRO, JESUS Centro: Universidad Carlos Iii

- Which are the research teams in the country • undertaking research in these fields?
- To which global, EU societal challenges are research groups contributing to?



ESCO categorization of ICT Job offers. Enriched Dashboard

- Job offers & BSc + MSc syllabus
- Job profiles automatically detected using topic analysis
- Jointly analyzed with other available metadata
- New metadata created using ML tools:
 - Gender aware detection
 - Average time that the job offers remain published
 - ESCO category





T08 - Desarrollo Web PHP - Symfony

T09 - Cloud Computing, Escalabilidad, Microservicios





Time interval de offer remained pul

• Resources mobilization: Is there a gap between supply and demand?

Province distribution





Type of degree required



Pole question!

(2) From what you have seen, do you think IntelComp tools will help you in your work?

Yes, matches our needs very well Yes, partially

Maybe, I am not sure yet

No

