

Flanders Research Information Space as a tool to monitor interdisciplinary research in Flanders

Bram Vancraeynest¹, Hoang-Son Pham¹, Hanne Poelmans¹, Sadia Vancauwenbergh¹

¹ECOOM-UHasselt
DSI-UHasselt

15th International Conference on Current Research Information Systems (CRIS2022), May 2022

- 1 Flanders Research Information Space (FRIS)
- 2 Measuring IDR
 - Organisational approach
 - Cognitive approach
- 3 Conclusions

Flanders Research Information Space (FRIS)¹

- regional CRIS of Flanders
- connected with the CRIS-systems of all Flemish universities and other knowledge institutions
- CERIF as exchange format
- used by the Flemish government for reports, analysis and statistics in the context of policy making and for monitoring trends in research and innovation.

¹www.researchportal.be/en

- 40.000 researchers
- 2000 research groups
- 50.000 projects
- 520.000 publications
- patents, datasets and research infrastructure

- Stimulating and measuring interdisciplinarity in research is one of the points of interest of the Flemish government

FRIS as tool for measuring interdisciplinarity

- Stimulating and measuring interdisciplinarity in research is one of the points of interest of the Flemish government
- Our research: measuring IDR (interdisciplinary research) in research projects using FRIS.

FRIS as tool for measuring interdisciplinarity

- Stimulating and measuring interdisciplinarity in research is one of the points of interest of the Flemish government
- Our research: measuring IDR (interdisciplinary research) in research projects using FRIS.
- Goal: Develop an indicator that measure IDR in projects on FRIS that uses all relevant information/data that a project on FRIS has.

Definition of IDR

Interdisciplinary research is a mode of research by teams or individuals that **integrates** information, data, techniques, tools, perspectives, concepts, and/or theories from **two or more disciplines** or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice.²

²Committee on Facilitating Interdisciplinary Research, Committee on Science, Engineering, and Public Policy (2004). Facilitating interdisciplinary research. National Academies. Washington: National Academy Press, p. 2.

- What exactly is integration?
 - Definition does not directly translates into a mathematical formula

- What exactly is integration?
 - Definition does not directly translates into a mathematical formula
- What information is relevant for IDR?
 - Researchers? Organisations? Disciplines? Title? Keywords? Abstract?
...

- What exactly is integration?
 - Definition does not directly translates into a mathematical formula
- What information is relevant for IDR?
 - Researchers? Organisations? Disciplines? Title? Keywords? Abstract?
...
- The notion of IDR depends on the choice of disciplines

Disciplines in FRIS

Disciplines in FRIS

- Flemish Research Discipline Standard (VODS)
 - 4 hierarchical levels

Disciplines in FRIS

- Flemish Research Discipline Standard (VODS)
 - 4 hierarchical levels
- Persons, organisations, projects have disciplines on FRIS
- Publications do not have disciplines on FRIS

Disciplines in FRIS

- Flemish Research Discipline Standard (VODS)
 - 4 hierarchical levels
- Persons, organisations, projects have disciplines on FRIS
- Publications do not have disciplines on FRIS

For IDR: level 2 VODS disciplines

- 42 disciplines
- Mathematics, physics, ..., arts

Disciplines in FRIS

- Flemish Research Discipline Standard (VODS)
 - 4 hierarchical levels
- Persons, organisations, projects have disciplines on FRIS
- Publications do not have disciplines on FRIS

For IDR: level 2 VODS disciplines

- 42 disciplines
- Mathematics, physics, ..., arts

Distance between disciplines

- Similarity based on the collaborations in projects
- E.g. $d(\text{math}, \text{physics}) = 0.2$ and $d(\text{math}, \text{arts}) = 0.86$

Projects

- Identifier, start date, end date, ...
- Title, abstract, keywords, ...
- Disciplines, participants, organisations, funding ...

Projects

- Identifier, start date, end date, ...
- Title, abstract, keywords, ...
- Disciplines, participants, organisations, funding ...

Blue: organisational approach

Red: cognitive approach

Persons

- Identifier, Gender, Keywords, Person names, Postal address ...
- Disciplines, Affiliations ...
- Projects
- Publications
- ...

Persons and cfPers elements in FRIS

Persons

- Identifier, Gender, Keywords, Person names, Postal address ...
- Disciplines, Affiliations ...
- Projects (+ co-participants)
- Publications (+ co-authors)
- ...

Blue: used to calculate 'true' disciplines of a researcher.

Persons and cfPers elements in FRIS

Persons

- Identifier, Gender, Keywords, Person names, Postal address ...
- Disciplines, Affiliations ...
- Projects (+ co-participants)
- Publications (+ co-authors)
- ...

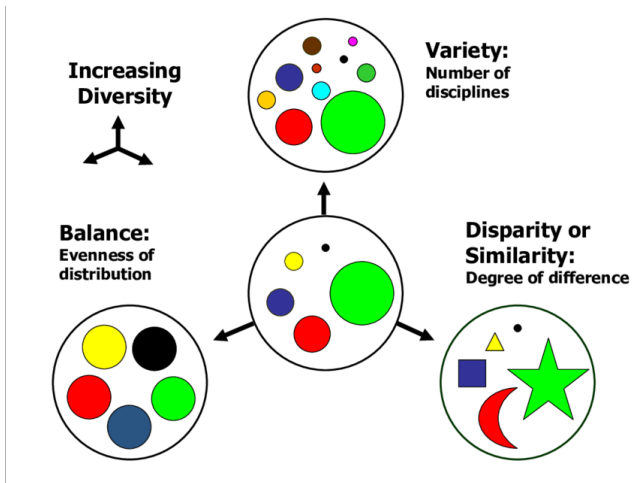
Blue: used to calculate 'true' disciplines of a researcher.

$$Dis(p) = w_1 \cdot D(profile) + w_2 \cdot D(affiliations) + w_3 \cdot D(projects) + w_4 \cdot D(coparticipants) + w_5 \cdot D(publications) + w_6 \cdot D(coauthors)$$

$$(w_1, w_2, w_3, w_4, w_5, w_6) = (0.35, 0.20, 0.20, 0.05, 0.15, 0.05)$$

- Organisational approaches
 - Diversity of researchers
 - Diversity of organisations
- Cognitive approaches
 - Diversity and network coherence of keywords
 - Diversity of topics through topic modeling

Rao-Stirling Diversity



Organisational approach

Assumption: a diverse team of researchers and organisations indicates a higher possibility of interdisciplinarity in a project.

Assumption: a diverse team of researchers and organisations indicates a higher possibility of interdisciplinarity in a project.

Measures:

- Diversity of researchers
- Diversity of organisations

- 1 Researchers in terms of their disciplines

1 Researchers in terms of their disciplines

- Jack (100% mathematics = $(1, 0, \dots, 0)$)
- Lisa (50% mathematics + 50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)
- Emma (50% mathematics +50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)

- 1 Researchers in terms of their disciplines
 - Jack (100% mathematics = $(1, 0, \dots, 0)$)
 - Lisa (50% mathematics + 50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)
 - Emma (50% mathematics +50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)
- 2 Distances between the researchers based on the distances of their respective disciplines

- 1 Researchers in terms of their disciplines
 - Jack (100% mathematics = $(1, 0, \dots, 0)$)
 - Lisa (50% mathematics + 50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)
 - Emma (50% mathematics +50% physics = $(\frac{1}{2}, \frac{1}{2}, 0, \dots, 0)$)
- 2 Distances between the researchers based on the distances of their respective disciplines
 - $d(L, E) = 0$ (they are the same in terms of disciplines)
 - $d(J, E) = d(J, L) = 0.5 \cdot d(\text{math}, \text{phys}) = 0.5 * 0.2 = 0.1$
- 3 Diversity of researchers

$$\frac{1}{3} \cdot \frac{1}{3} d(J, L) + \frac{1}{3} \cdot \frac{1}{3} d(J, E) + \frac{1}{3} \cdot \frac{1}{3} d(L, E) = 0.022$$

- Approaches based on content of project
- Content of a project = Title + Keywords + Abstract
- Assumption: “keywords” embedded in content from different disciplines indicate possibility of interdisciplinarity in a project

- Approaches based on content of project
- Content of a project = Title + Keywords + Abstract
- Assumption: “keywords” embedded in content from different disciplines indicate possibility of interdisciplinarity in a project

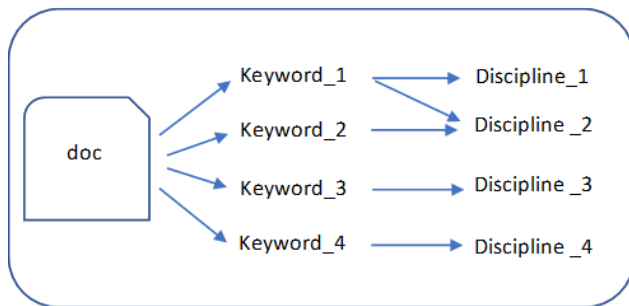
Our proposed approaches

- Diversity and network coherence of keywords
- Topic modeling

Step 1: Keyword extraction based on title, abstract and keywords

Diversity of keywords

Step 1: Keyword extraction based on title, abstract and keywords



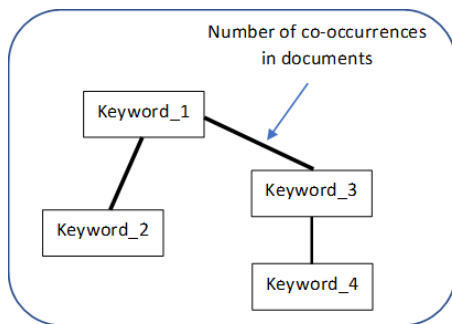
Diversity of keywords = diversity of disciplines that keywords belong to

Coherence network of keywords

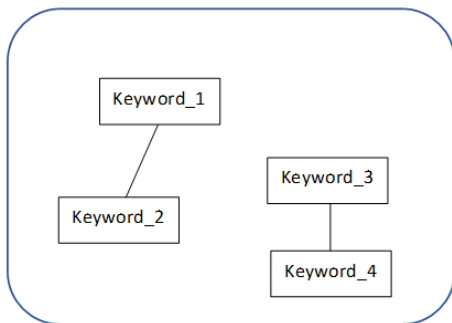
Step 1: Keyword extraction based on title, abstract and keywords

Coherence network of keywords

Step 1: Keyword extraction based on title, abstract and keywords



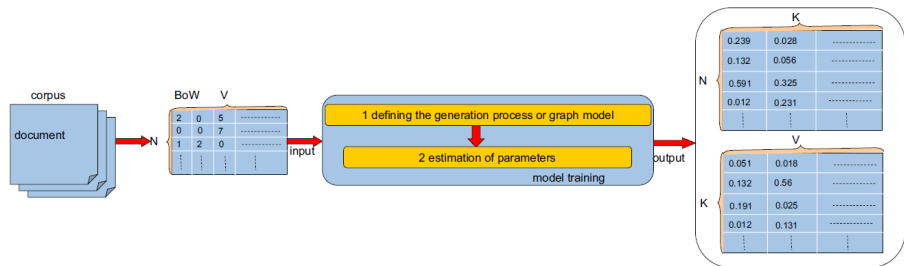
High coherence network



Low coherence network

Low network coherence indicates that the keywords are less correlated, and that there is a higher possibility for knowledge integration

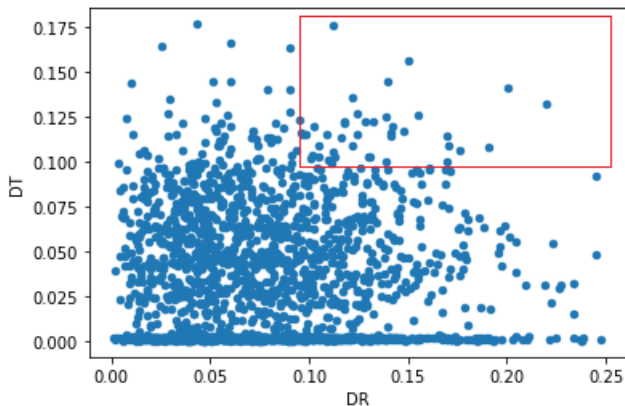
Topic modeling framework



- Unsupervised topic modeling (e.g, Latent Dirichlet Allocation (LDA))
- Supervised topic modeling (e.g., Labelled LDA)

Combination of Organisational and Cognitive approach

Diversity of researchers (DR) and Diversity of topics (DT) of 2283 projects



High DR and high DT indicate potential IDR

- FRIS is a CRIS with several purposes including monitoring research
- Our research: how can we use FRIS to measure IDR in project
- IDR is complex and cannot be captured in an exact mathematical definition
- Our own version of IDR based on the project data available on FRIS combining two facets of IDR:
 - Organisational: participants, organisations
 - Cognitive: disciplines, title, abstract, keywords,

- FRIS offers lot of data where we that we can use to measure IDR

Strengths, shortcomings and remarks

- FRIS offers lot of data where we that we can use to measure IDR
- Our methods depend on the data quality
 - short abstracts, 'duplicate' projects, ...

- FRIS offers lot of data where we that we can use to measure IDR
- Our methods depend on the data quality
 - short abstracts, 'duplicate' projects, ...
- Possible useful additions
 - project proposals
 - publication disciplines

Thank you for your attention!