

Recommender Systems using Pennant Diagrams in Digital Libraries

NKOS Workshop

London, 2014-09-12

Zeljko Carevic and Philipp Mayr

firstname.lastname@gesis.org

Introduction

- Recommender Systems are an established way to lead users to related content.
- Often the users demand a detailed view on the connection between a document and its connections.
 - Who's work is related to the current document / topic?
 - What other descriptors are related to the current document / topic?
- What's missing is the distance between the current document and the recommendations.
- One way of showing the distance is using so called Pennant Diagrams.

Pennant Diagrams

- Method to visualize the relevance / relatedness of a given seed to Documents / Authors / Descriptors in a Scatter Plot.
- Pennant Diagrams combine methods from:
 - Relevance Theory
 - Information Retrieval
 - Bibliometrics



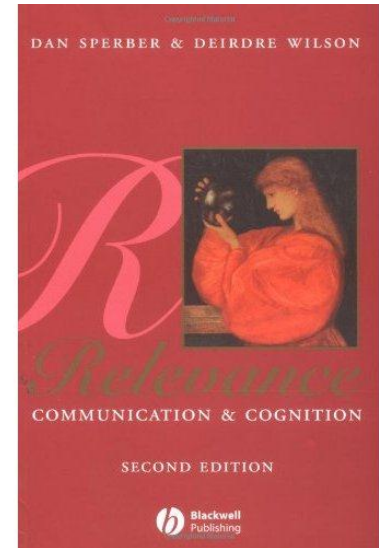
Created by Howard D. White
Drexel University

Pennant Diagrams

Relevance Theory



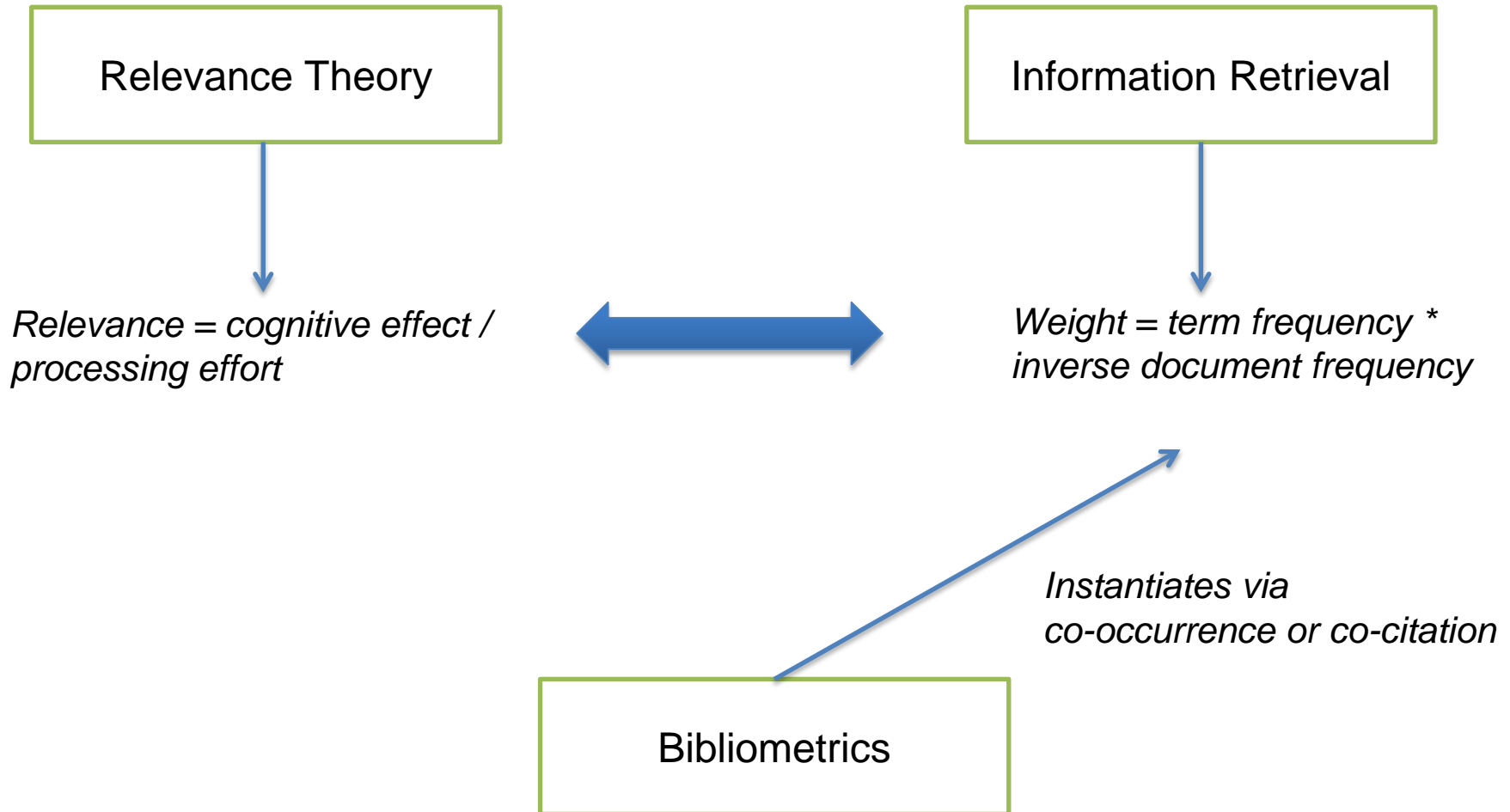
Relevance = cognitive effect / processing effort



Cognitive effect: The **greater** the cognitive effect the **more relevant** it becomes

Processing effort: The **less** processing effort is necessary the **more relevant** it becomes

Pennant Diagrams



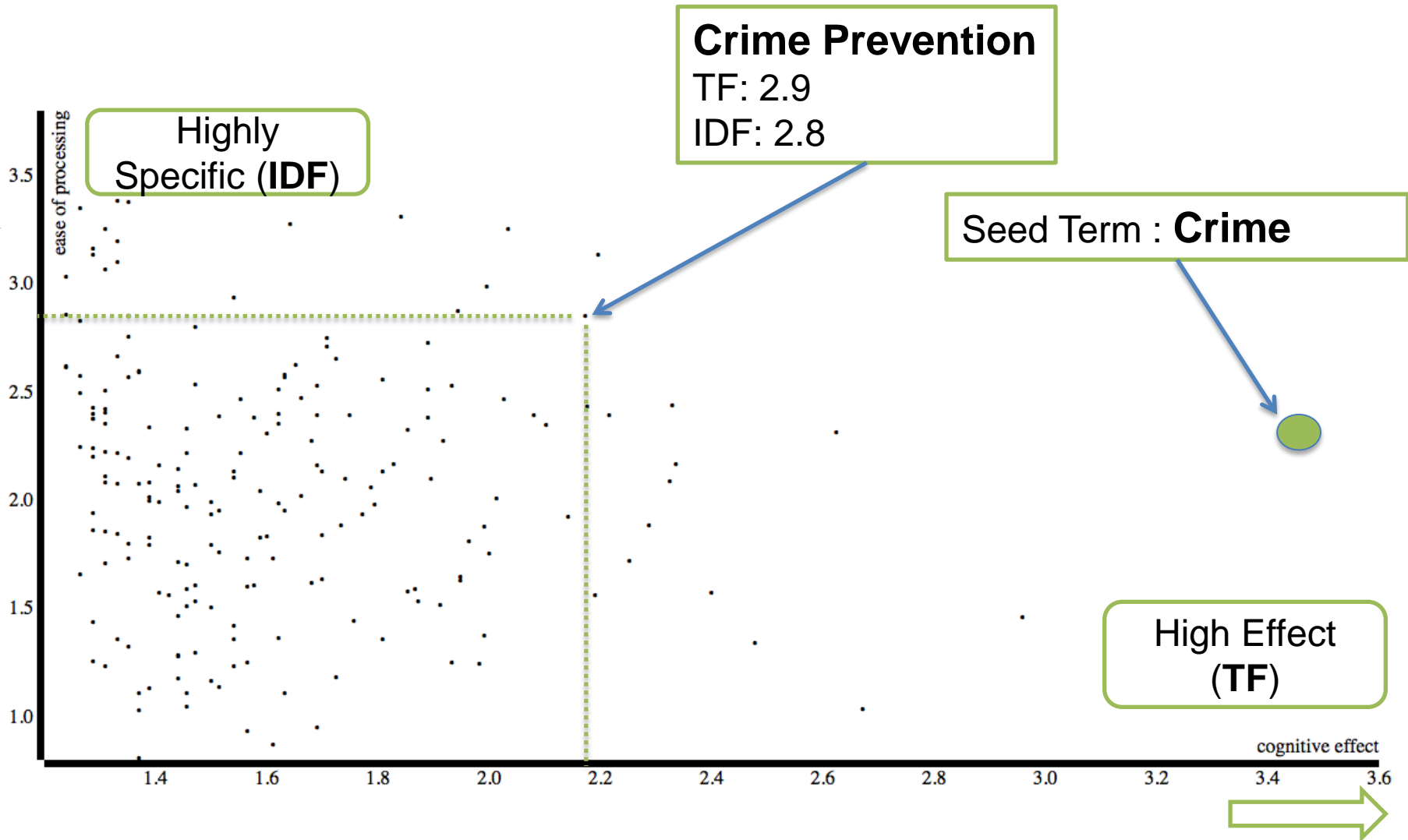
Calculating TF / IDF

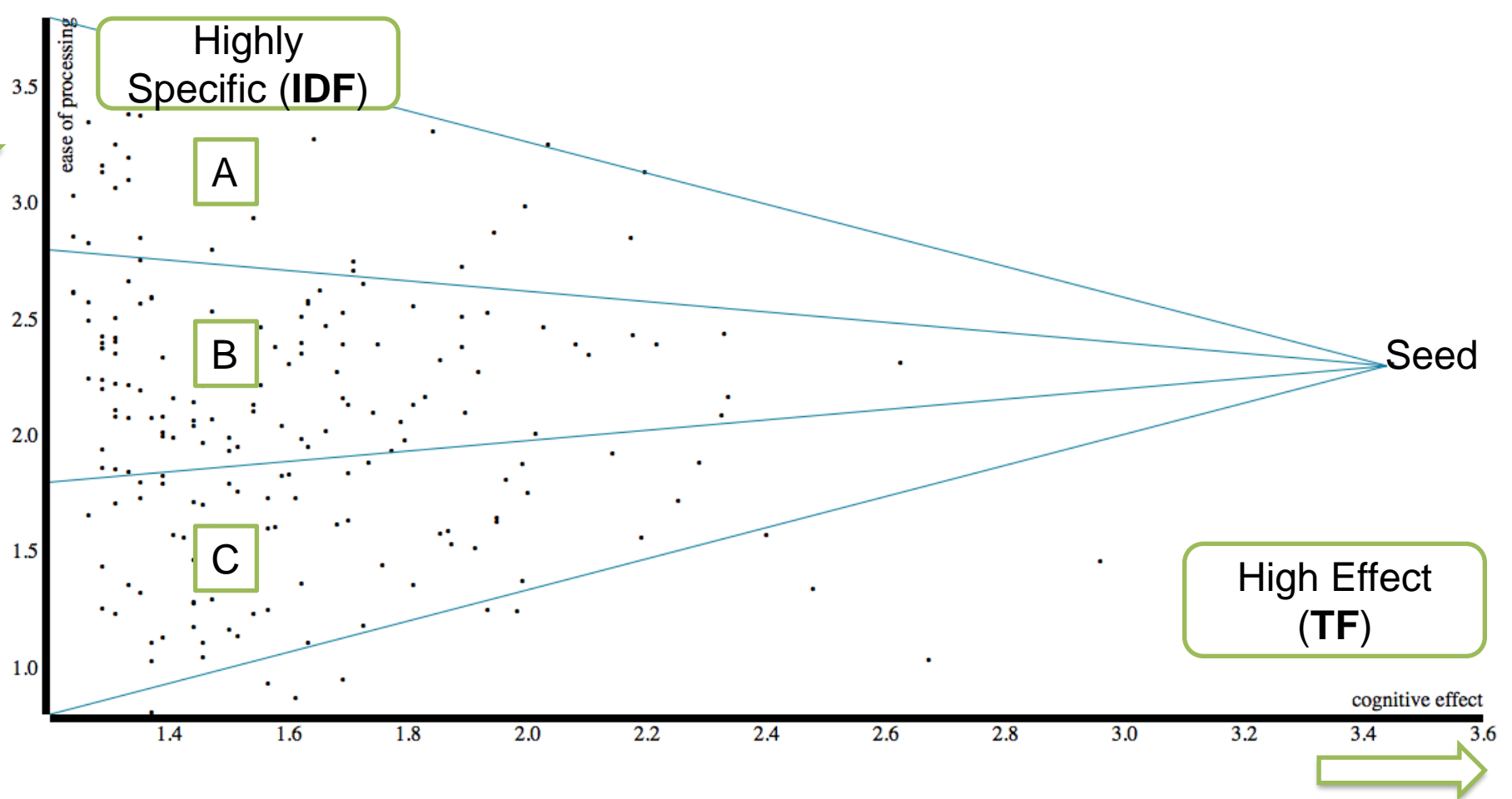
IR - TF*IDF ranking

- Starts with a query term
- **tf** = Term frequency in current doc
- **df** = Number of docs query term appears in
- **TF*IDF** = similarity between **doc** and **query term**

Co-Occurrence - TF*IDF ranking

- Start with a seed term
- **tf** = Number of times a term co-occur with seed
- **df** = Number of times a term occur overall
- **TF*IDF** = similarity between **doc** and the **seed**





Use Case

- Support researchers in:
 - Lead researchers into new directions
 - Discovering new Descriptors
 - Discovering new Authors
 - Allow explorative searching
 - Recommender System

Sowiport

- **Sowiport**: A digital library for the social sciences
- Containing about **8. mio** records with metadata and links to full-text
- Documents **contain citation information** and **descriptors**
- Using **Apache Solr** as Search Index

The screenshot displays the Sowiprot website interface. At the top, the 'gesis' logo is on the left, and a 'Login' button is on the right. Below this is the 'sowiport' logo with the tagline 'The online portal for the social sciences'. A navigation bar contains 'Home', 'Search result', 'Parents' duties, children's...', and 'Description'. The main content area is divided into two columns. The left column, titled 'Similar items', lists several related documents with their titles and years. The right column features a book icon and the title 'Parents' duties, children's debts : the limits of policy intervention'. Below the title, it provides publication details: 'Published: Aldershot: Arena, 1995. VIII, 188 p.', 'Topics: Child care; Great Britain; Aged; Care; Parent and child [Law]; Family policy; Family; Children; Old persons', and 'Language: Englisch (EN)'. The document type is listed as 'Buch'. At the bottom of the right column, there are tabs for 'Description' and 'Comments', and a metadata section showing 'Database: DZA-GEROLIT, ID: dza-gerolit-18837907X' and 'Copyright: DZA'. A footer at the very bottom contains the text '© GESIS | Imprint | About Sowiprot | Partners | Team | Feedback'.

Implementation using Java Script

1. Start with a seed term: **Crime**

Lookup „crime“ in Solr **including Facets**

Lookup each Facet in Solr



Descriptor	Tf	Df
Crime	35.270	35.270
Violence	1767	
Police	1688	

Implementation using Java Script

1. Start with a seed term: **Crime**

Lookup „crime“ in Solr **including Facets**

Lookup each Facet in Solr



Descriptor	Tf	Df
Crime	35.270	35.270
Violence	1767	46.517
Police	1688	27.245

Violence co-occurs 1767 times with **Crime**
Violence occurs 46.517 times in sowiport

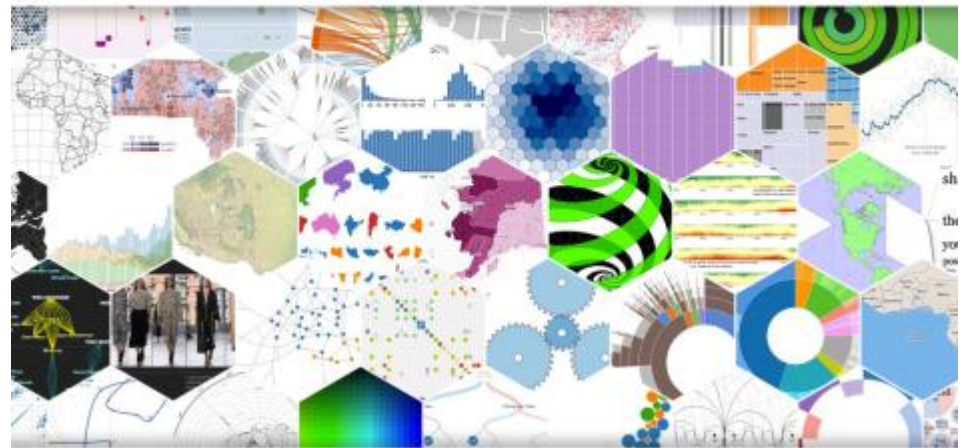
D3 Framework for Visualizing

- Java Script framework to visualize large datasets
- Instantiated using JSON representation of co-occurring descriptors

```
{ tf=1767, df=46517,
  name="Violence" }
```

- Visualization separated from model-building

Data-Driven Documents



Demo

Discussion and future work

- Preliminary results of implementing Pennant Diagrams in a digital library.
- **Future Work:**
 - Implement Pennant Diagrams with Co-Citation Data
 - Integrate visualization in Sowiport
 - Evaluate with Users
 - Filter Descriptors (Black List)
- **Questions:**
 - How to display a huge amount of terms on one pennant?
 - Are the chosen sectors appropriate?
 - How to evaluate the diagram?