

## **Linking of Bibliographic and Research Data in Social Sciences**

Dr. Christian Hänger  
(Mannheim University Library)  
Brigitte Mathiak (GESIS)

## Problem Statement

Task: Write a paper about the political opinions of elites, based on Mannheim research

Library

Books and more Catalogue Inter-Library Loan

elite mannheim political opinions search

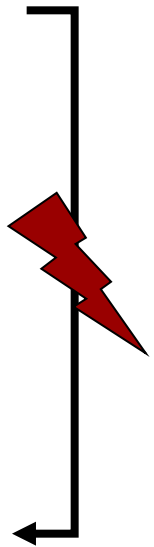
All items that contain my query words anywhere in the record

0 Results for the library's holdings

→ no appropriate literature available

Gesis

| ZA1138: Mannheim Elite Study 1968   |                           |
|---|---------------------------|
| Bibliographic Citation Content Methodology Data & Documents Errata & Versions Further Remarks |                           |
| Study No.   | ZA1138                    |
| Title   | Mannheim Elite Study 1968 |



## Problem Statement


Task: Compare the results of a specific book with your own work about the political opinions of elites

Books and more | Catalog | Inter-Library Loan

Oberschicht und politisches System der Bundesrepublik Deutschland  [Advanced Search](#)

All items ▾ that contain my query words ▾ anywhere in the record ▾

Library

 **Oberschicht und politisches System der Bundesrepublik Deutschland : soziale Mobilität und Karrieremuster von 800 Inhabern von Spitzenpositionen der westdt. Gesellschaft**  
Edo Enke  
Bern [u.a.] : Lang 1974  
**Available ; several locations**

[Locations](#) [Details](#) [Request](#) [More](#)

 no link to research data

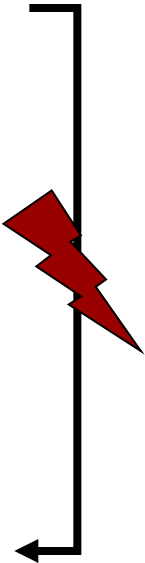
Gesis

**ZA1138: Mannheim Elite Study 1968**

[Bibliographic Citation](#) [Content](#) [Methodology](#) [Data & Documents](#) [Errata & Versions](#) [Further Remarks](#)

Publications

- Enke, Edo:  
[Oberschicht und politisches System der Bundesrepublik Deutschland.](#)  
Frankfurt: Lang 1974.



## Realization

1. Analyze the GESIS metadata format for studies
2. Transform the format into compatible one for the RDS
3. Load the data into RDS
4. Create links between publications and research data

## Exemplary Search

Books & more **GESIS**

employment and wages

search

that contain my query words ▼

anywhere in the record ▼


1 Results for **MAN GESIS**

sorted by: Relevance ▼

1



resdata

 **Employment and Wages in German Industries between 1888 and 1954**  
**Beschäftigung und Löhne der deutschen Industriewirtschaft 1888-1954**  
DIVO, Frankfurt [Data Coll.]  
GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

[Details](#)

1 Results for **MAN GESIS**

sorted by: Relevance ▼

# Exemplary Study in Primo

Books & more **GESIS**

employment and wages

search

[advanced search](#)

that contain my query words ▼

anywhere in the record ▼

**Employment and Wages** in German Industries between 1888 and 1954

Beschäftigung und Löhne der deutschen Industriegewirtschaft 1888-1954

DIVO, Frankfurt [Data Coll.]

GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

**Details**

Title: **Employment and Wages** in German Industries between 1888 and 1954  
Beschäftigung und Löhne der deutschen Industriegewirtschaft 1888-1954

Author: DIVO, Frankfurt [Data Coll.]  
GESIS - Leibniz-Institut für Sozialwissenschaften [Publ. Agent]

Subjects: Historische Sozialforschung ;  
Historical Social Research

Description: Methodik: Mündliche Befragung mit standardisiertem Fragebogen  
Methodik: Oral survey with standardized questionnaire  
Beschreibung: In der vorliegenden Studie wird versucht, Beschäftigungs- und Lohnentwicklung in Deutschland auf Grund der Unterlagen der Unfallversicherungsstatistik zu ermitteln, da die oben erwähnten Bedenken diesem Material gegenüber entfallen. Die Angaben über die Zahl der

Publisher: GESIS - Leibniz-Institut für Sozialwissenschaften

Creation Date: 2010-04-13

**Links**[> Research Data Source](#)

# Link Literature and Research Data


Standorte **Details** Bestand Weitere Services

|                   |  |
|-------------------|--|
| Titel:            | Altersbild und Altersvorsorge der Arbeiter und Angestellten      |
| Autor:            | Ludwig von Friedeburg ;<br>Friedrich Weltz                       |
| Schlagwörter:     | Arbeiter / Altersversorgung ;<br>Angestellter / Altersversorgung |
| Verknüpfte Titel: | Frankfurter Beiträge zur Soziologie : Sonderheft ; 1             |
| Verlag:           | Frankfurt am Main : Europ. Verlagsanst.                          |
| Erscheinungsjahr: | 1958   |

**Links**

> [Find Primary Research Data](#)

➔ link to research data

- 1  [Image of the Old and Provision for Old Age for Workers and Employees](#)  
**Altersbild und Altersvorsorge der Arbeiter und Angestellten**  
 DIVO, Frankfurt [Data Coll.]  
 GESIS - Leibniz-Institut für Sozialwissenschaften 2010-04-13

## Create links between publications and research data

### Problem:

- Searching for study names in publications is insufficient
  - Study names may be commonly used nouns like „Gesundheit“ (health) or „Wirtschaft und Ethik“ (economy and ethics)
  - Studies are often referenced using abbreviations and synonyms

### Solution:

- Use Machine Learning methods to identify patterns in which dataset references typically occur
- Use Bootstrapping to overcome sparse data problem

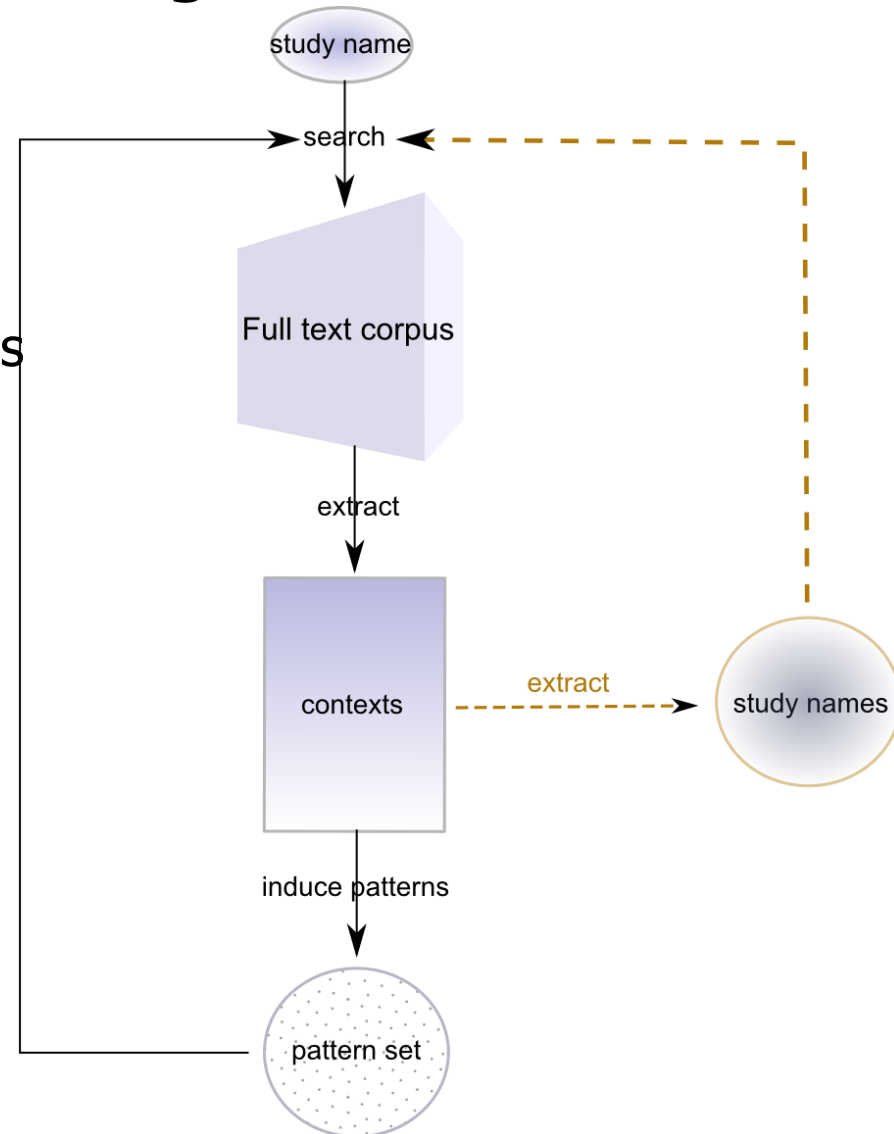


## Bootstrapping: Pattern Induction

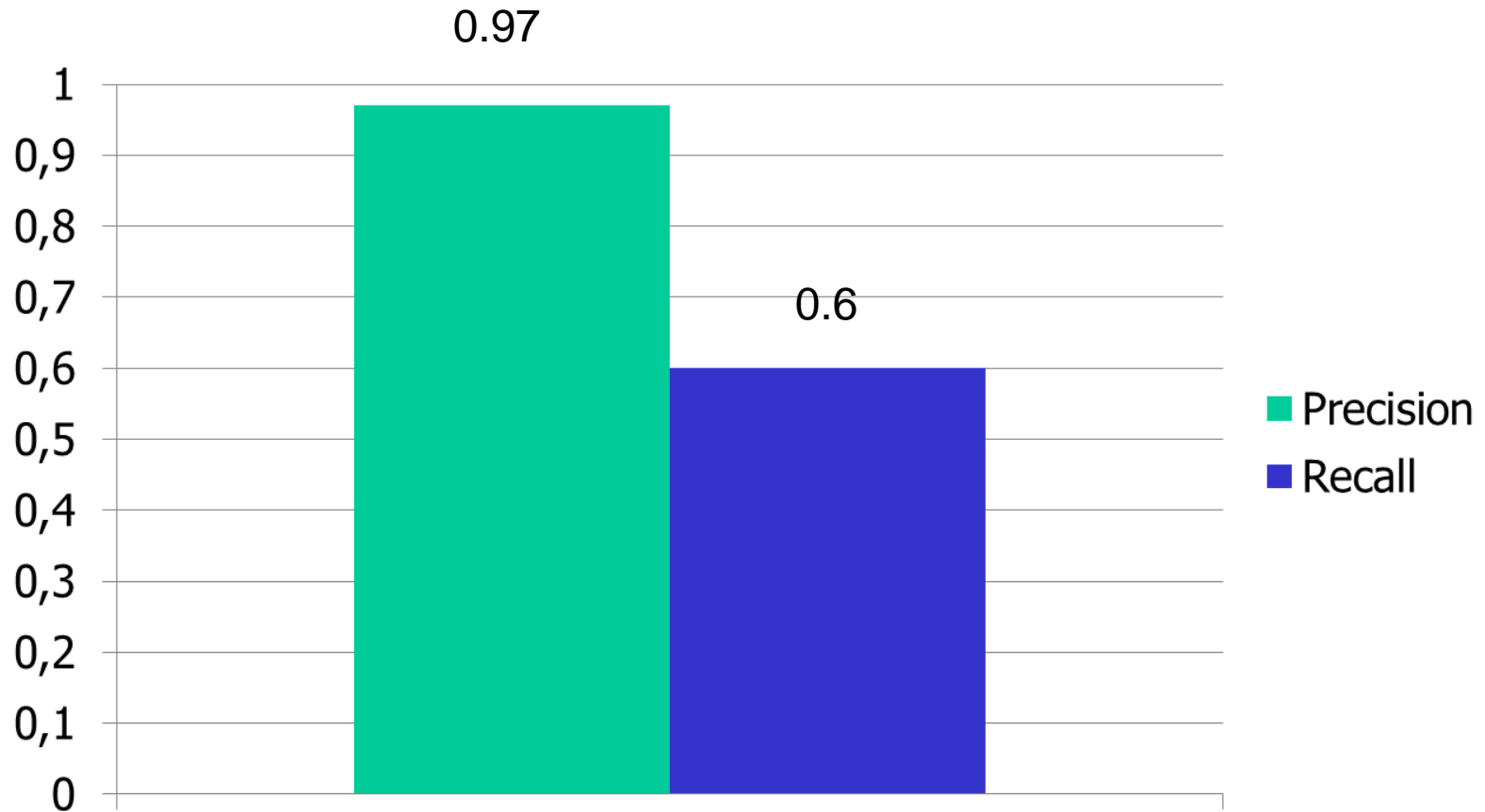
1. Induce patterns from small seed training set
2. Find new instances by applying the patterns on the test set
3. Add new instances to the training set
4. Go to 1 using the extended training set

## Find references to datasets using Pattern Induction

- Search for term „ALLBUS“ in text corpus
- Extract contexts
- Induce patterns
- Search the corpus for these patterns
- Extract new study names
- Search corpus for new study names
- Generate new contexts and patterns
- If no new study names can be found: terminate



- SSOAR documents with keyword „empirisch-quantitativ“ (empirical-quantitative)
- DGS subcorpus (German, n=259)
- Preprocessing
  - conversion Pdf → Txt
  - automatic correction
  - Removal of bibliographies



Thank you for your attention!