Automatic extraction of citation data in a distributed digital library

José M. Barrueco
Thomas Krichel
Automatic extraction of citation data in a distributed digital library

CitEc

CitEc:

A system to reference linking documents available on the RePEc digital library

Developed at the Universidad de Valencia
Automatic extraction of citation data in a distributed digital library
Reference linking

B
C
D
E
F
G

A

The scientific literature
Automatic extraction of citation data in a distributed digital library

Related work

• CrossRef (PILA)
• CiteSeer (NEC Research Department)
• OpCit: Open Citation Project (JISC + NSF)
• CERN
• CitEc (UPV)
  - (heterogeneous documents + distributed environment)
Automatic extraction of citation data in a distributed digital library
The RePEc data set

RePEc (Research Papers in Economics):

- non-commercial
- distributed (+200 institutions)
- archives vs services
- largest academic digital library:
  - 149,127 documents
  - 61,406 of them are downloadable
  - 984,582 Abstract views
  - 112,420 Downloaded documents
Automatic extraction of citation data in a distributed digital library
CitEc as data provider

INPUT

RePEc

Paper and Article Templates

OUTPUT

RePEc

Citation Templates

+ Paper and Article Templates
Automatic extraction of citation data in a distributed digital library

The extraction process

READING
- Metadata
- Full Text (PDF)

PARSING
- PDF
- ASCII
- References

LINKING
- Reference
- Linking
- Citation Template

Knowledge Base

RePEc Knowledge Base

Communication

Automatic extraction of citation data in a distributed digital library

The extraction process

READING
- Metadata
- Full Text (PDF)

PARSING
- PDF
- ASCII
- References

LINKING
- Reference
- Linking
- Citation Template

Knowledge Base

RePEc Knowledge Base

Communication
Automatic extraction of citation data in a distributed digital library
Implementation

Providing the latest research results since 1993

Search: [ ] tips: [ ] title= fiscal or author=levine

Working Papers Series | Journals | Authors | JEL Classification
--- | --- | --- | ---


Technology Innovations and the Volatility of Output: An International Perspective

Christian Zimmermann
zimmermann.christian@uqam.ca
(Center for Research on Economic Fluctuations and Employment, UQAM (web site)

Author(s) registered at HoPEc:
CHRISTIAN ZIMMERMANN

Cahiers de recherche CREFE / CREFE Working Papers 34 / CREFE, Université du Québec à Montréal (web site)
(RePEc:cre:crefwp:34)

Related papers by JEL classification:
E32: Business Fluctuations, Cycles
F41: Open Economy Macroeconomics

Download this document

Application/PostScript (Main text)
Application/PDF (Main text)

See this paper’s references

Cited by:
- RePEc:cre:crefwp:33
- RePEc:cre:crefwp:37
- RePEc:cre:crefwp:38
- RePEc:cre:crefwp:45
- RePEc:cre:crefwp:131
- RePEc:sie:sieaen:69-98
Automatic extraction of citation data in a distributed digital library
Implementation

My References:


Automatic extraction of citation data in a distributed digital library

Results

Processed documents:

- Analizados: 48%
- Error en conversión: 12%
- Sin referencias: 6%
- Error en identificación de referencias: 3%
- Formato no aceptado: 7%
- No en inglés: 24%
Automatic extraction of citation data in a distributed digital library

Conclusions

• System to automate the process of reference linking in a distributed digital library
• The citation data has been fully integrated with the existing RePEc services
• We need better conversion tools
• We need to improve the parsing algorithm
Automatic extraction of citation data in a distributed digital library

The End

Questions, comments, even money are accepted