

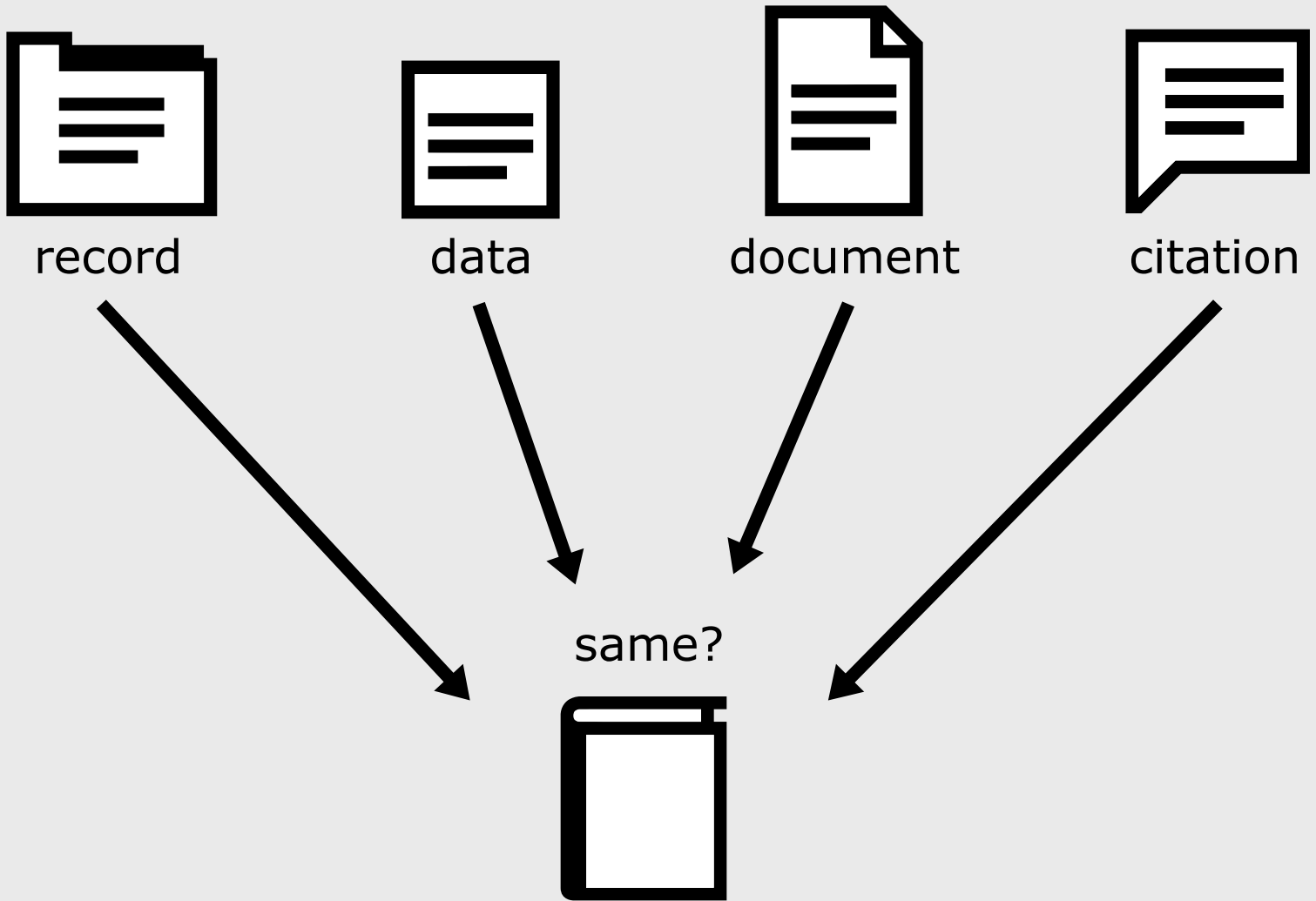
Mapping Bibliographic Records with
Bibliographic Hash Keys

Jakob Voß

Verbundzentrale des GBV (VZG)

11. Internationales Symposium für
Informationswissenschaft (ISI 2009)

the problem



examples

```
004A      $00262025388
011@      $a2003
021A      $aA @history of online information
           services, 1963 - 1976
028A      $dCharles P.$aBourne$9383311691
028B/01   $dTrudi Bellardo$aHahn$9271366931
033A      $pCambridge, Mass. [u.a.]$nMIT Press
034D      $aXVI, 493 S
```

Bourne, C.P.; Hahn, T.B. (2003): A History of Online Information Services, 1963-1976. – Cambridge, Mass.; London: MIT Press.

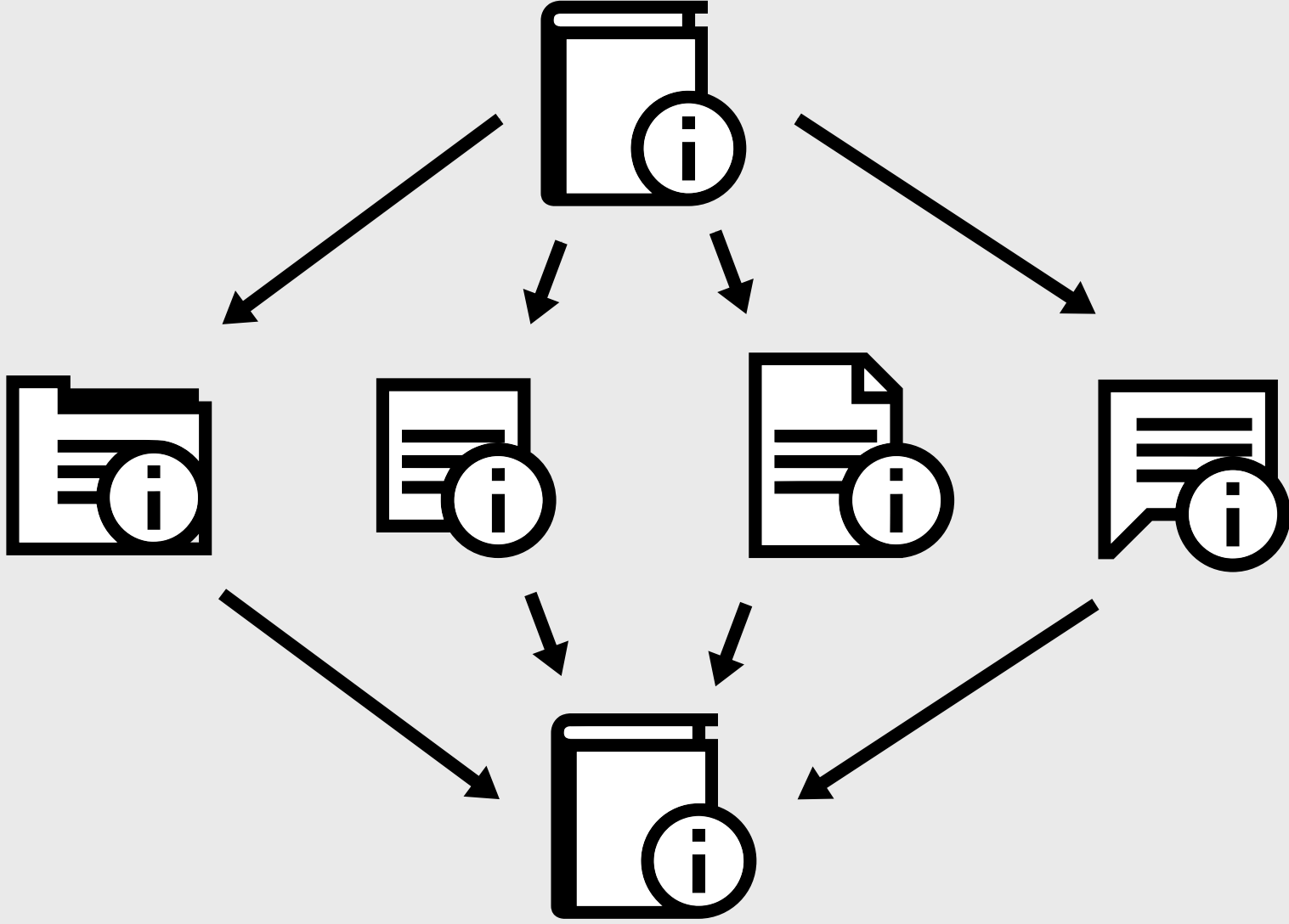
```
@book{Bourne2003,
  title={A History of Online Information Services, 1963-1976},
  author={T.B. Hahn and C.P. Bourne},
  year={2003},
```

Charles P. Bourne and Trudi Bellardo Hahn, A History of Online Information Services, 1963-1976 (Cambridge: MIT Press, 2003),

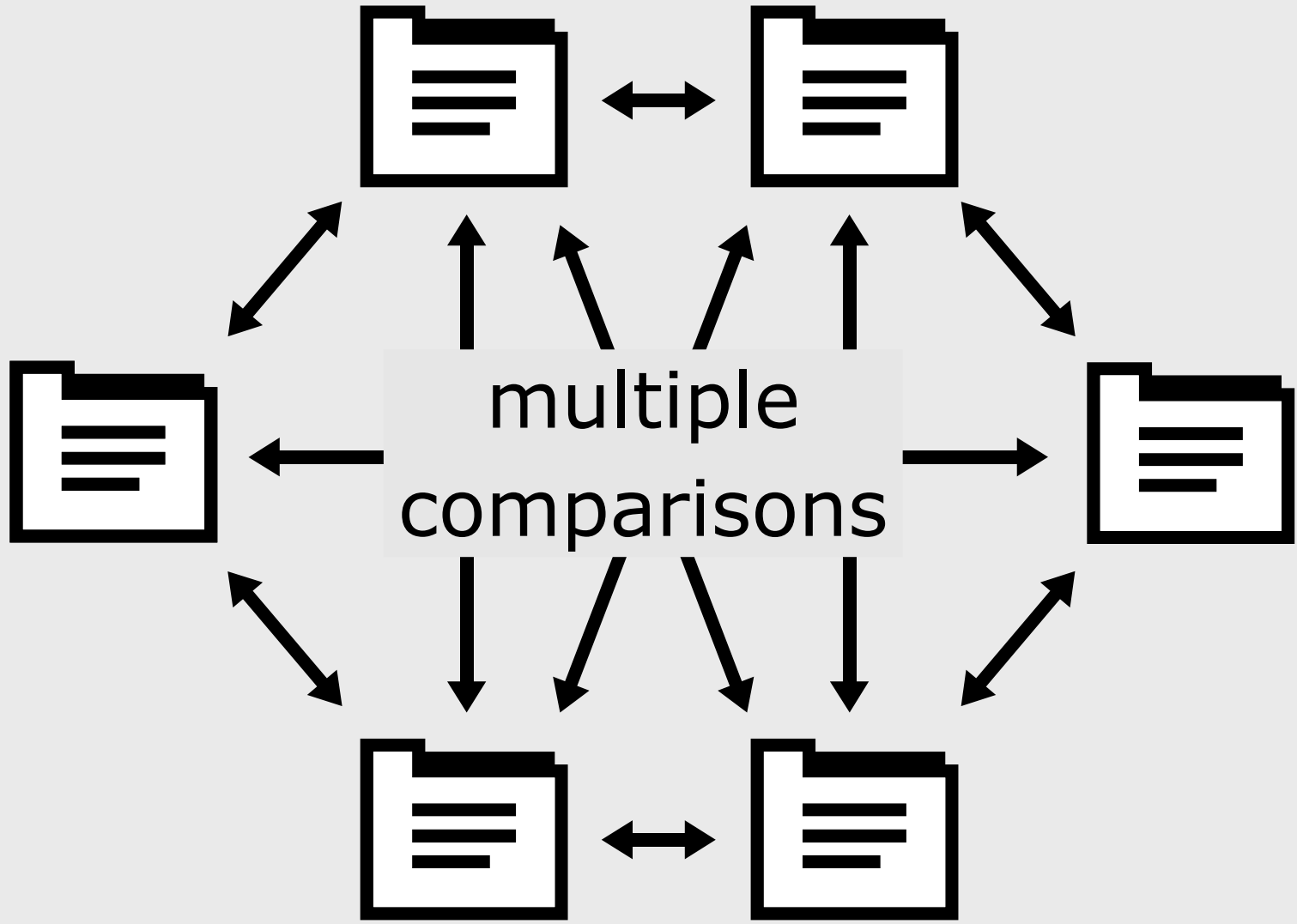
solutions

- (a) persistent identifiers
- (b) duplicate detection
- (c) **fingerprints**

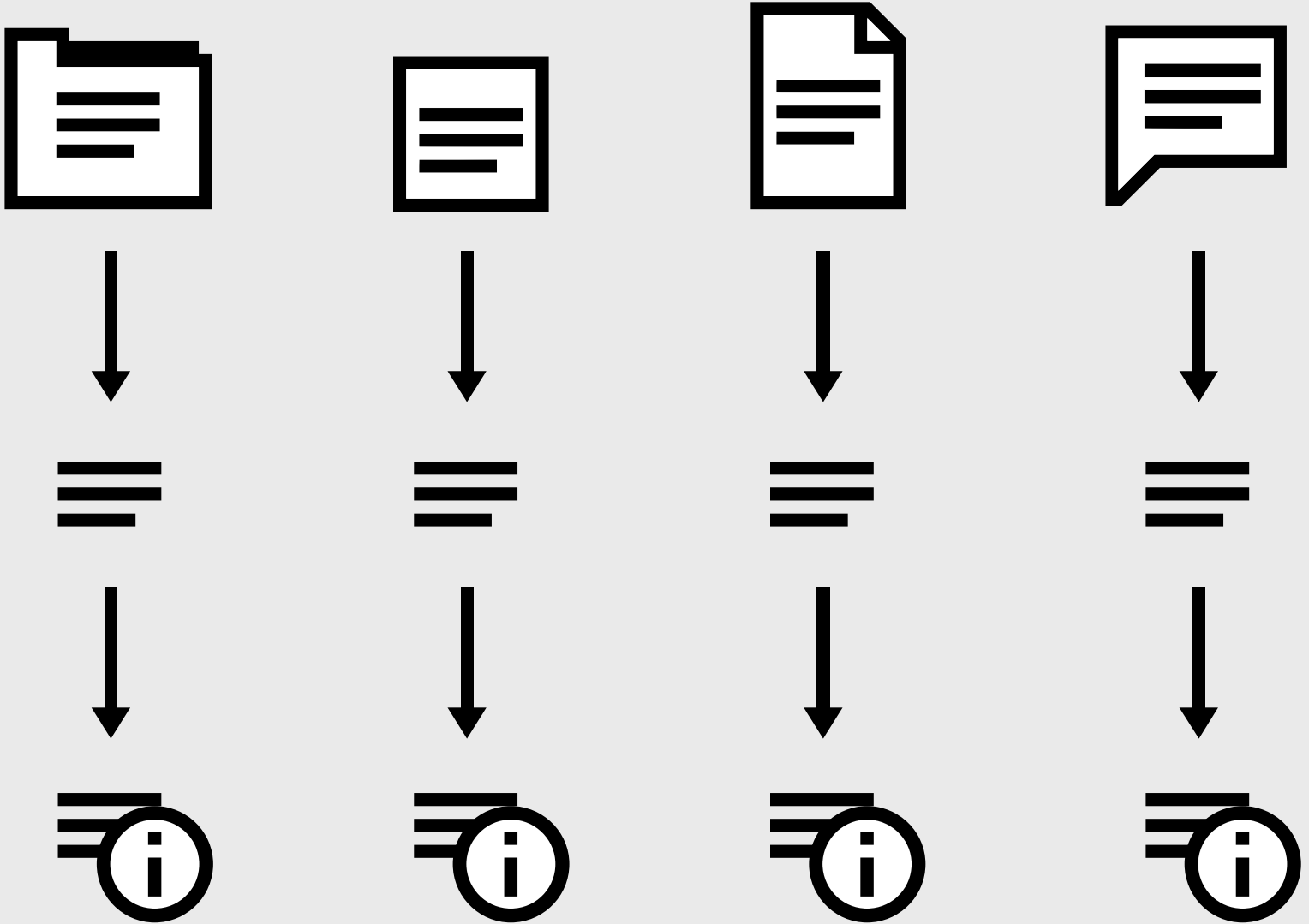
persistent identifiers



duplicate detection



fingerprints (bibkey)



specification of bibkey

1. selection of fields
2. normalization
3. concatenating
4. prefix and hashing

selection of fields

```

004A      $00262025388
011@      $a200 3
021A      $aA @history of online information
           services, 1963 - 1976
028A      $dCharles P. $aBourne $9383311691
028B/01   $dTrudi Bellard o$aHahn $9271366931
033A      $pCambridge, Mass. [u.a.]$nMIT Press
034D      $-XVT 402 3

```

**Bourne, C.P.; Hahn,
of Online Information
– Cambridge, Mass.;**

**authors/editors
title
year**

```

@book{Bourne2003, title=
{A History of Online Information Services, 1963-1976 },
author={T.B. Hahn and C.P. Bourne },
year={2003},

```

**Charles P. Bourne and Trudi Bellardo Hahn,
A History of Online Information Services,
1963-1976 (Cambridge: MIT Press, 2003),**

normalization

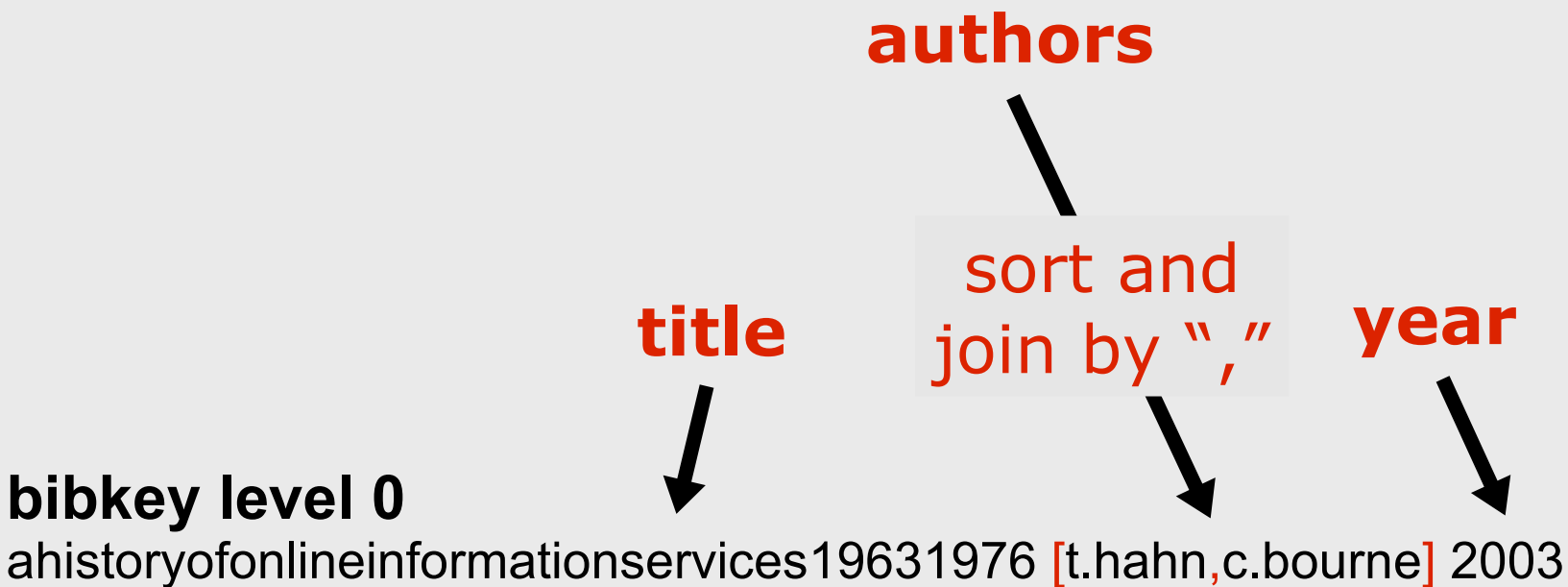
lowercase, letters, digits

Charles P. Bourne → c.bourne

Trudi Bellardo Hahn → t.hahn

unicode normalization form
compatibility composition (NFKC)

composing the hash



prefix "1" and MD5 hash

bibkey level 1

14ed100f75dd4459cfeeb272bdbbc2d1e7

usage in BibSonomy

bibkey

The screenshot shows a Mozilla Firefox browser window displaying a BibSonomy page. The address bar contains the URL `http://www.bibsonomy.org/bibtex://1acfec953843b168e61e2e167e29b4c3d`, which is circled in red and labeled "bibkey". The page title is "BibSonomy :: bibtex :: Detecting Communities via Simu". The page content includes a list of publications and a tag cloud. A red bracket groups the first three publications, labeled "independent bibliographic records". Another red bracket groups the tag cloud, labeled "aggregated tag cloud".

publications (5) [RSS](#) [BibTeX](#) [RDF](#) [more](#)

previous | 1 | next [edit](#) | [pick](#) | [unpick](#)

Detecting Communities via Simultaneous Clustering of Graphs and Folksonomies
Akshay Java and Anupam Joshi and Tim Finin *Proceedings of the Tenth Workshop on Web Mining and Web Usage Analysis (WebKDD)* (2008)
to 2008 clustering community detection folksonomy graph social-media by annaba on 2009-01-14 10:25:57
[pick](#) | [copy](#) | [URL](#) | [BibTeX](#) | [OpenURL](#)

Detecting Communities via Simultaneous Clustering of Graphs and Folksonomies
Akshay Java and Anupam Joshi and Tim Finin *WebKDD 2008 Workshop on Web Mining and Web Usage Analysis* (2008)
to clustering detecting folksonomies graphs by nesty.nov on 2009-01-10 16:44:24
[pick](#) | [copy](#) | [BibTeX](#) | [OpenURL](#)

Detecting Communities via Simultaneous Clustering of Graphs and Folksonomies
Akshay Java and Anupam Joshi and Tim Finin *WebKDD 2008 Workshop on Web Mining and Web Usage Analysis* (2008)
to clustering community detection by folke on 2009-01-07 07:06:51
[pick](#) | [copy](#) | [BibTeX](#) | [OpenURL](#)

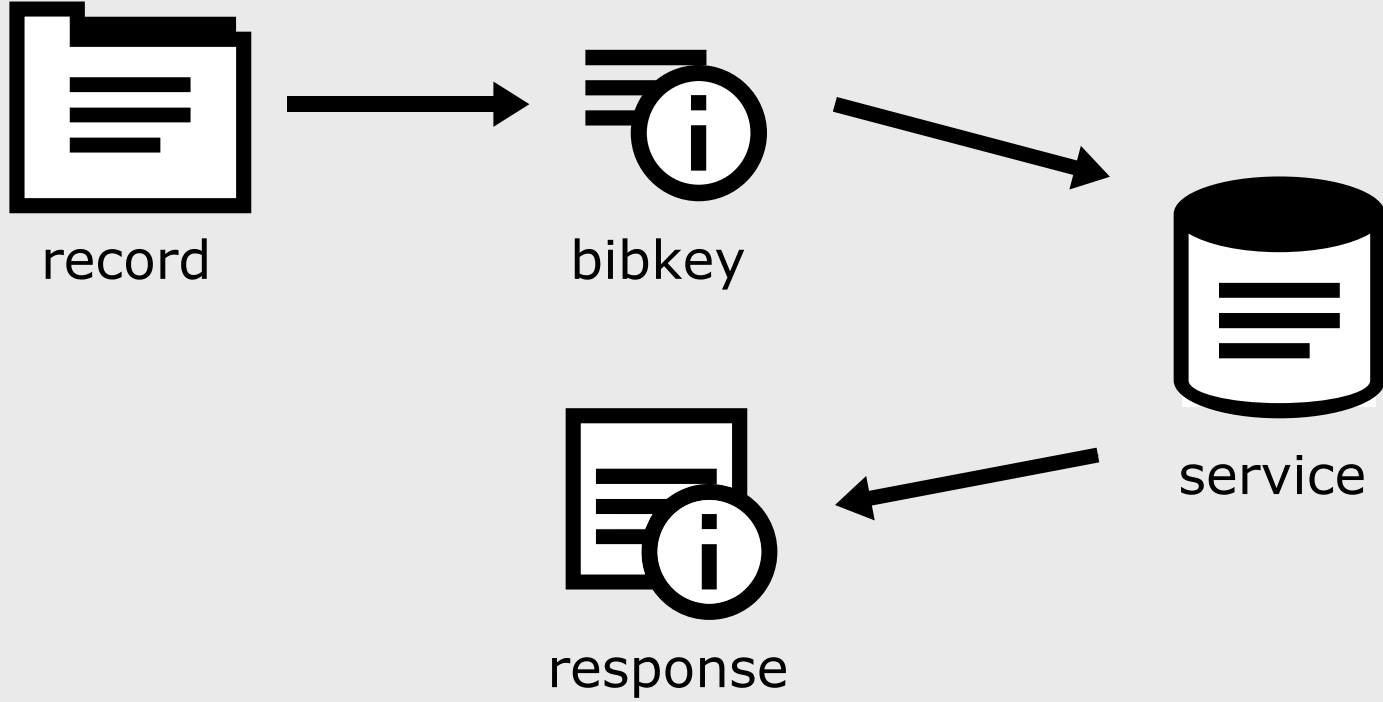
filter:

tags

2008 clustering clustering
community detecting
detection folksonomies
folksonomy graph graphs
social-media toread

Fertig

usage in web services



challenges

- determining
 - all authors
 - the full title
 - the year
- problematic works
 - no authors
 - standard titles
(“Introduction”, “News”...)
- unicode normalization

www.gbv.de/wikis/cls/bibkey



icons based on

PICOL