

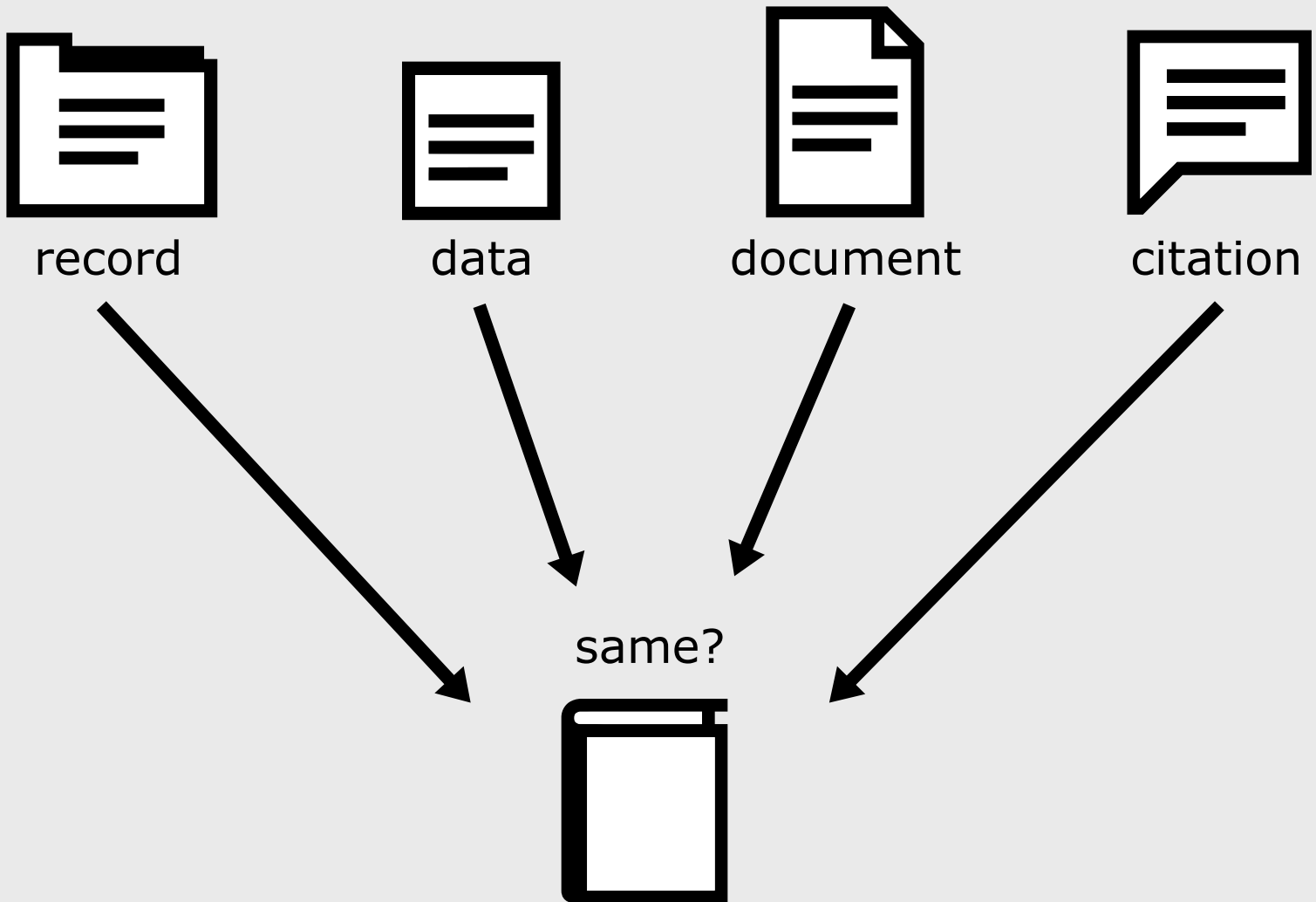
Mapping Bibliographic Records with
Bibliographic Hash Keys

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},

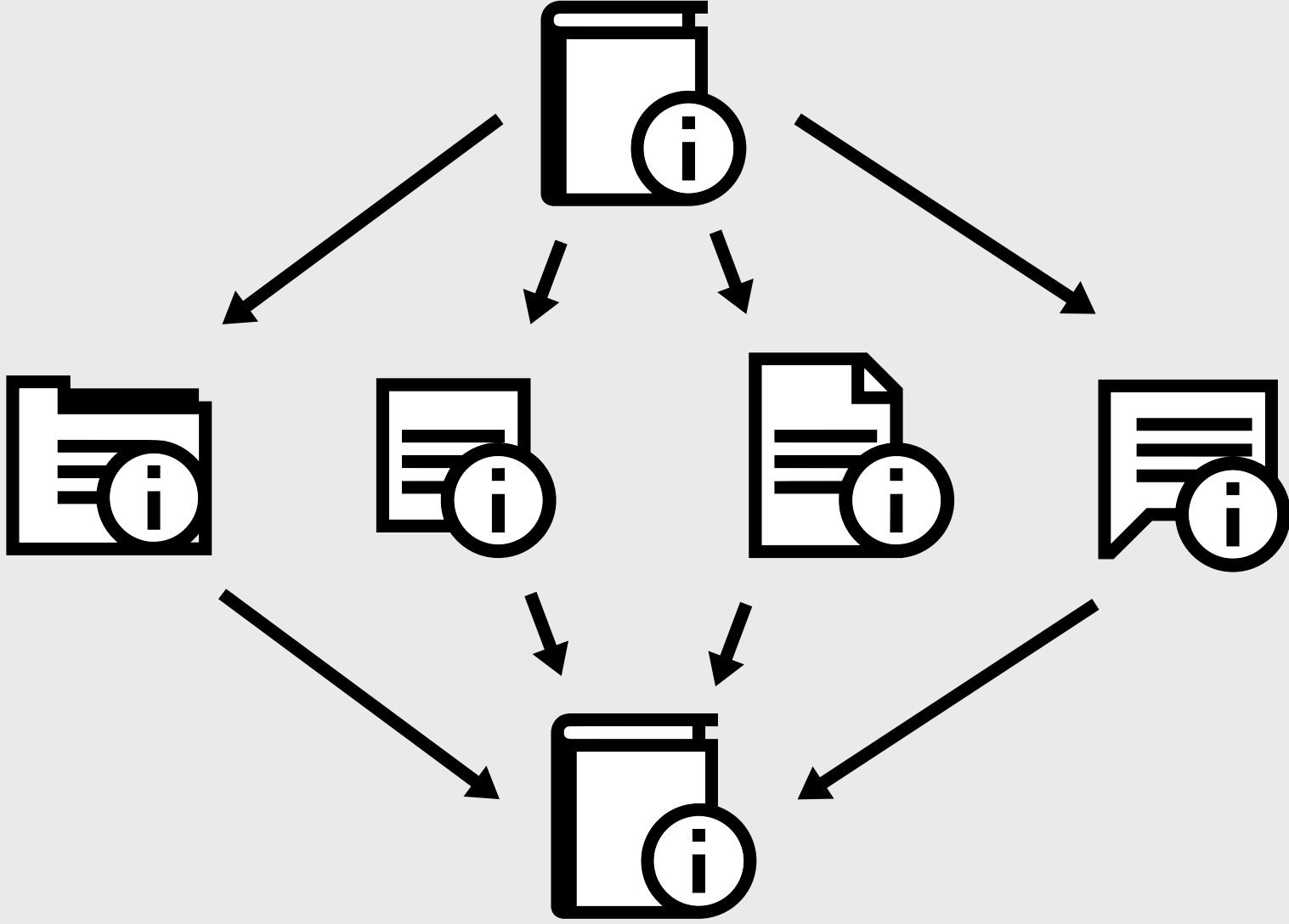
```

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

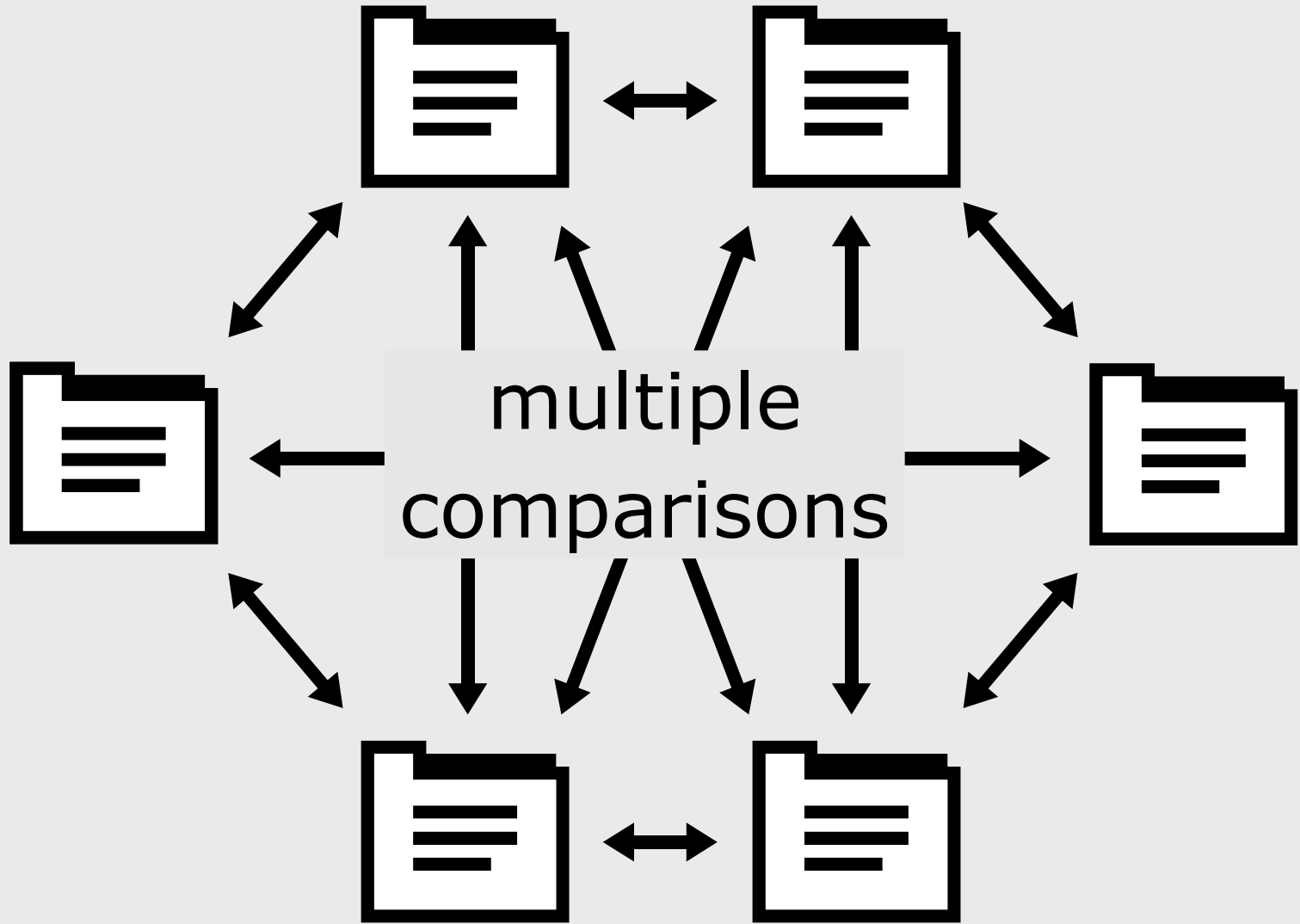
solutions

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

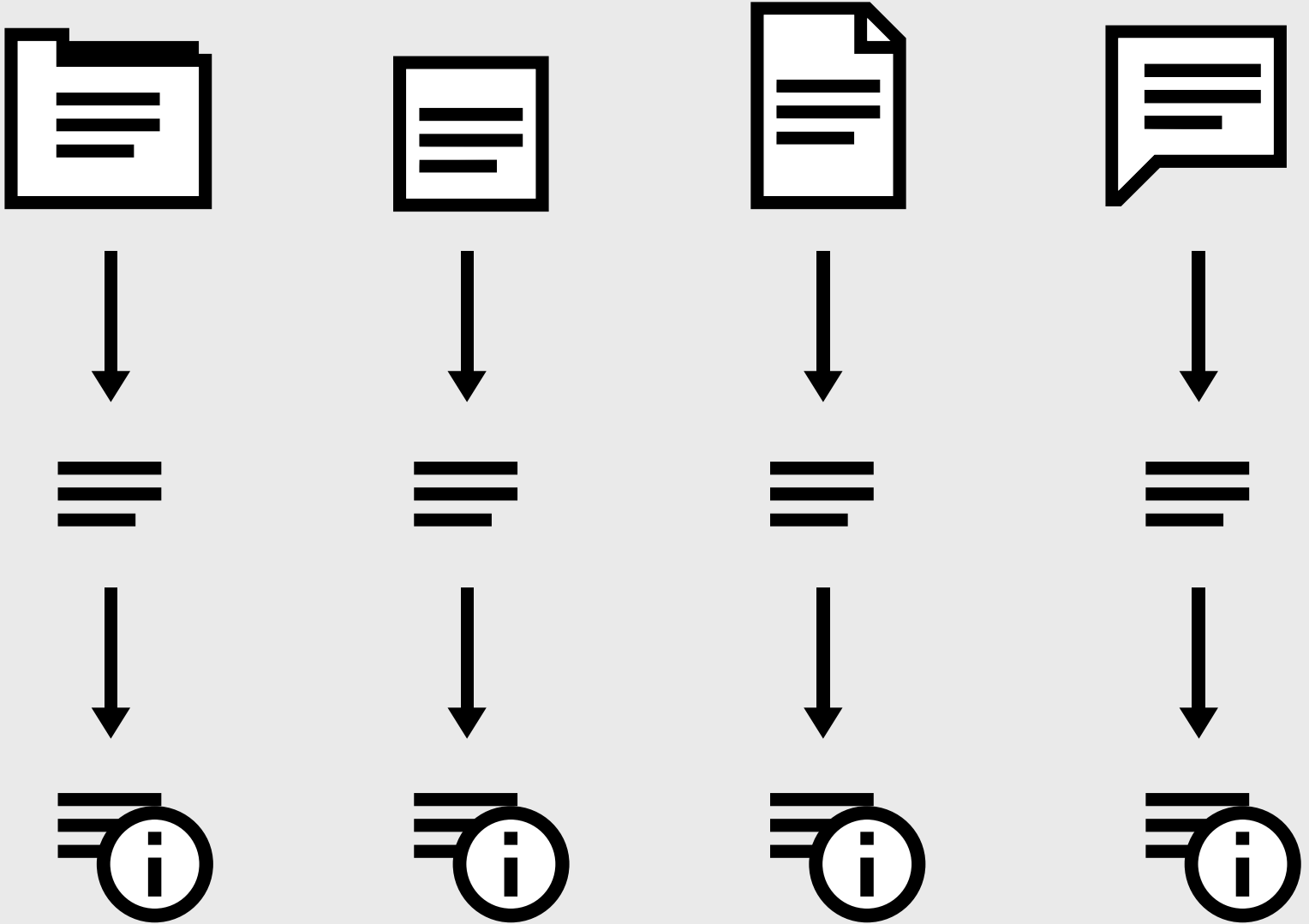
persistent identifiers



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

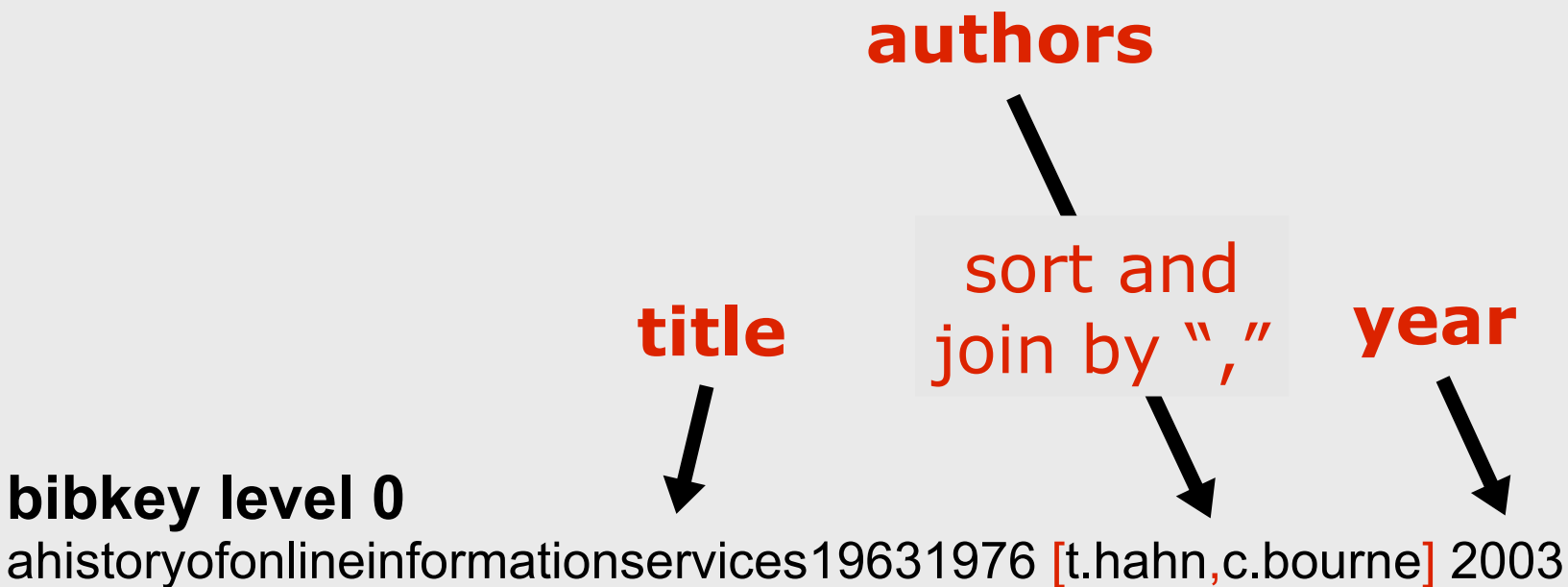
lowercase, letters, digits

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

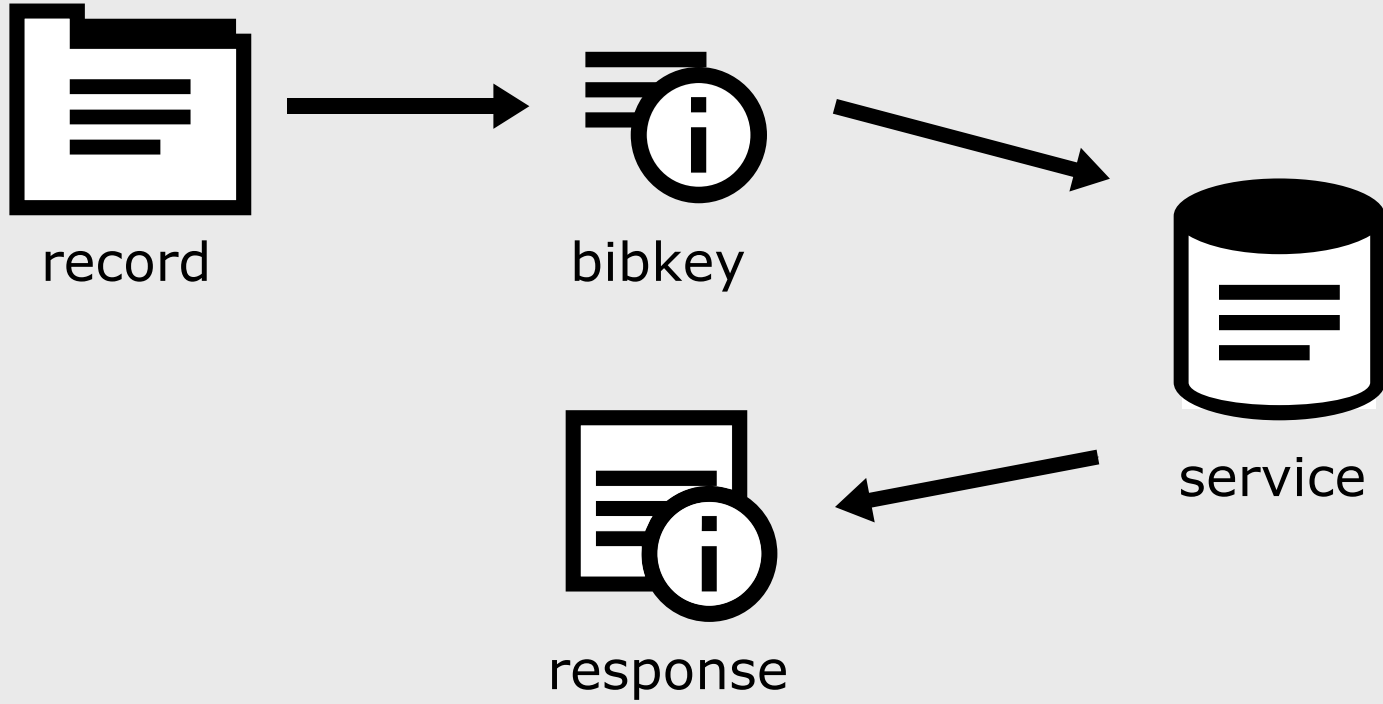
14ed100f75dd4459cfeeb272bdbbc2d1e7

usage in BibSonomy

bibkey

The screenshot shows the BibSonomy website interface. The browser address bar contains the URL `http://www.bibsonomy.org/bibtex/1/1acfec953843b168e61e2e167e29b4c3d`, which is circled in red and labeled "bibkey". The page title is "BibSonomy :: bibtex :: Detecting Communities via Simu". The main content area displays a list of publications, with the first one titled "Detecting Communities via Simultaneous Clustering of Graphs and Folksonomies" by Akshay Java and Anupam Joshi and Tim Finin. A red bracket groups these three publication entries, labeled "independent bibliographic records". To the right of the list is a tag cloud, also labeled "aggregated tag cloud". The tag cloud contains various tags such as "2008", "clustering", "community", "detection", "folksonomy", "graph", "graphs", "social-media", and "toread". The browser window also shows a "logged in as hotho" status and a "zotero" extension icon in the bottom right corner.

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