Using external ranking tools

May 8, 2012
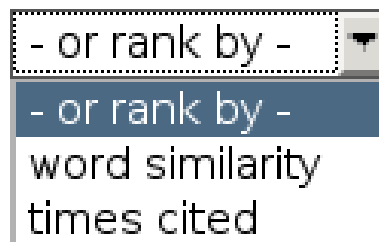
Patrick Glauner
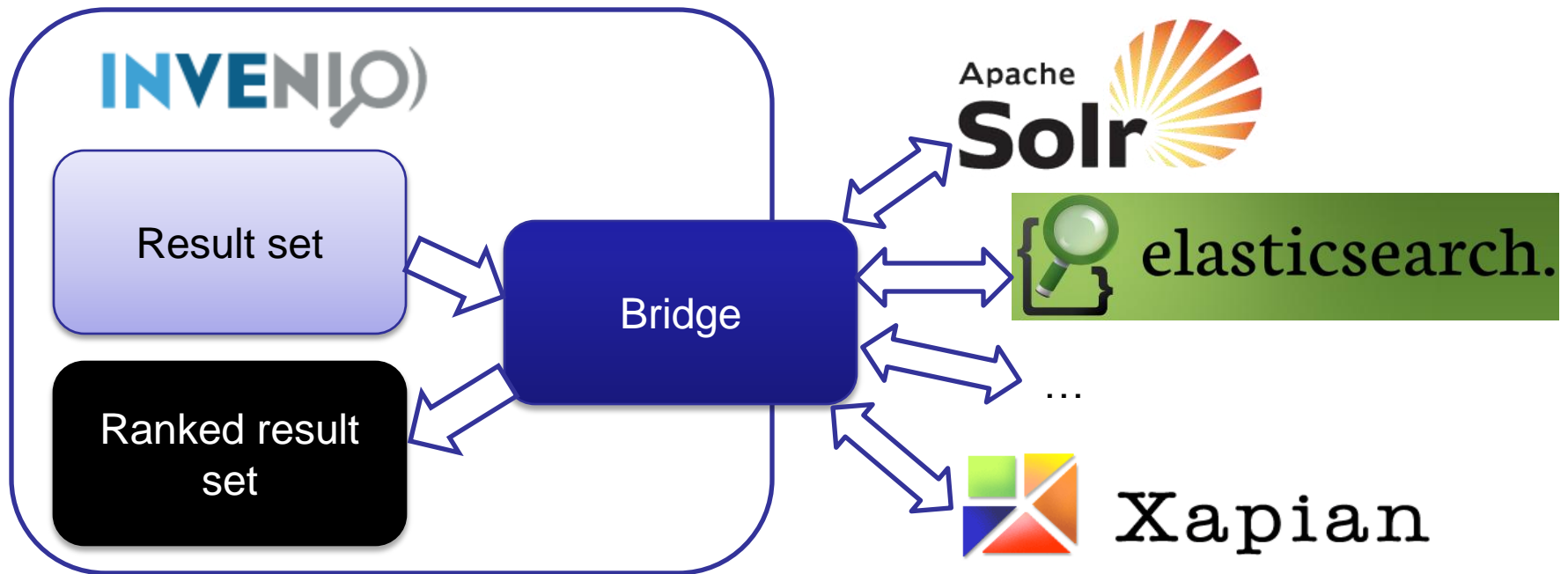patrick.oliver.glauner@cern.ch

- Available ranking methods:
  - **Word similarity**
  - Citation count

- Current word similarity ranking implementation:
  - **Meta-data only**
  - Written in **Python**
  - **Maintenance** and **further development consume development resources** that could be spent on digital library features
  - **Scalability**, designed for ~2 million records

- A generic **bridge** between Invenio and external information retrieval systems

- Invenio shall dispatch word similarity ranking tasks to the external system and retrieve its results

Apache Solr and **elasticsearch.**

– Written in Java and Lucene-based

– HTTP access

	Apache Solr	elasticsearch.	Xapian
Platform independence	+	+	+
Documentation	+	-	0
Python integration	+	-	+
Data format definition	0	+	-
Configuration	0	-	+
Word similarity ranking	+	+	+

Xapian

– Written in C++

– Direct access to local database in program code

Progress and Outreach

- Apache **Solr** integration is currently being implemented
- Will be ready end of June including scalability measurements

- Future priorities will depend on the community feedback
- Possible future features:
  – Xapian adapter
  – Further ranking methods
  – External search
  – …