VISUALLY MINING SCIENTIFIC COMMUNITIES

BSc project by Silas Berger, supervised by Leonel Merino
E-Mail: silas.berger@students.unibe.ch
What problem we want to solve

- Valuable information is hidden in amounts of papers too large for humans to read
- Visualizations augments human understanding of data
What we are building on (EggShell)

- Can import PDFs as XML extract title and contributors and model retrieved data
- First visualization exists, but focus is on data extraction and (visual) assessment of precision

```model := EggShell modelPapersInFolder: 'examplePapers'
using: #ESXMLPipeline.
visualization := ESEventVisualizationExample
new
visualization setModel: model.
visualization open.`
Project outlook
Current work

- Query language (OCL-like approach)
- Data model expansion
- Work out sample queries to assess quality of data model and query language

```
(ExtendedEggShell papers select: [ :paper | paper authors includes: 'L. Merino' ]) flatCollect: [ :paper | paper authors ].
(ExtendedEggShell venue: #VISSOFT allAuthors) collect: [ :author | author affiliation ] sortByOccurrences: #DESC.
(ExtendedEggShell papers) lookFor: ('Roassal' 'communities' 'agile') collect: [ :paper | paper title ].
```
VISUALLY MINING SCIENTIFIC COMMUNITIES

• Based on EggShell
• Adapting the data model of scientific communities
• Language for querying the model
• Visualization of results that encourage exploration