Collaborative process maturing support by mining activity streams

iKnow 2015
Prof. Dr. René Peinl
Graz, 21.10.2015
Agenda

1. Use case / context
2. Background: established research streams
3. Process mining – the SCHub way
4. Live demonstration of parts of the solution
5. Conclusion and outlook
**Goal**: Establishing an integrated infrastructure for effective support of team collaboration, esp. for **knowledge intensive tasks** and regionally distributed employees

- direct support for knowledge and business processes
- From a user’s perspective, a unified intranet with continuous support for working tasks without breaches in the workflow should arise.

**Solution**: Integration of Open Source Software from the areas portal, document management (DMS), groupware and business process management (BPM)
Goal for SCHub process maturing support

- Allows for **collaboratively enhancing document-centric processes**
- Use **Web 2.0** feature like commenting, rating, tagging in BPM
- Use information available in **activity streams** to learn about workflows
- Support **weakly structured processes** with software
- Use **open standards** and available **open source software** as much as possible
Focus of the solution

Schmidt, A., Hinkelmann, K., Ley, T., Lindstaedt, S., Maier, R., Riss, U.: Conceptual foundations for a service-oriented knowledge and learning architecture: Supporting content, process and ontology maturing.
Background: Adaptive Case Management (ACM)

- ACM does not force strict workflows
- **Suggests situation-specific actions** that might be required
- Users can **adopt** suggested tasks, but can also **adapt** them
- No strict separation between design-time and runtime
- Complements Business Process Management (BPM)
- Object Management Group (OMG) Standards:
  - Case Management Model and Notation (CMMN, May 2014)
  - Business Process Model and Notation (BPMN v.2.0, January 2011)
- Camunda supports CMMN since version 7.2 (November 2014)
Example: Oracle Adaptive Case Management

Tasks available to me at this state of case

Guidance on which Task may be more valuable to perform

What are other users/roles doing to this case

Virtual Folder

Activity flow for selected task
Background: BPM research streams

• **Subject-oriented BPM**
  – Communication-oriented
  – Very easy for end-users
  – Maybe too simple for knowledge workers

• **Social BPM**
  – Bring Web 2.0 participative approaches to BPM
  – Potential benefits: increased transparency and knowledge sharing
  – Drawbacks: possibly lower quality process models, difficult to evaluate
Background: process mining and maturing

• Process Mining
  – Discovery type: event logs to process models
  – Conformance checking type: compare process model with log
  – Enhancement type: extend existing models with data from logs

• Process Maturing
  – Weakly structured process might result from missing knowledge
  – recording of the users’ activities => creation & maturing of processes
  – IT support for weakly structured processes is key to link organizational and personal knowledge work
Example: KISSmir research prototype (Witschel et al. 2010)

Own approach

• User activities from all systems are collected in Apache Shindig (OpenSocial) and stored in neo4j (graph database)
• Document creation and publishing timestamp as a frame for the context
• Document type, title and headlines are used as semantic context
• Initial case creation
  – Suggest tasks and milestones
• Case enhancements
  – Suggest task patterns / sequences
Task recommendation algorithm explained

- **t₀**: document created, document edited, document annotated, meta data edited, document reviewed, document v 1.0 published
- **t₁**:

**User group**

- **Similiarity**
- **Wiki pages**
- **Blog**
- **Keyword search**
- Referenced content per activity
- Aggregated task candidates

**Shindig Activities**

- 20/10/15 - 09:31: Bärbel Bitte hat Berge, Gletscher, Nebel... gepostet (Liferay Blogs)
- 17/09/15 - 11:32: Bärbel Bitte hat Tag "grün" zu Landschaft2 hinzugefügt
- 17/09/15 - 11:19: Bärbel Bitte hat Tags (telle, nummer) zu textfile hinzugefügt
- 09/09/15 - 10:19: Bärbel Bitte hat Java mit 5 Sternen bewertet (Liferay Wikis)
- 02/09/15 - 12:32: Bärbel Bitte hat Meeting in Calendar aktualisiert (Open-Xchange)
- 02/09/15 - 12:32: Bärbel Bitte wird an Meeting in Calendar teilnehmen (Open-Xchange)
- 02/09/15 - 12:31: Bärbel Bitte wird nicht an Meeting in Calendar teilnehmen (Open-Xchange)
- 02/09/15 - 12:27: Bärbel Bitte hat Meeting in Calendar aktualisiert (Open-Xchange)
Content similarity

- One German and one English corpus for testing
- Two news sites each: heise.de, golem.de, cnn.com, bbc.co.uk
- Document consists of selected articles from one source
- Text files with news contents from both sources of the language as reference
- Text similarity is calculated for the whole document and for single chapters

- Term frequency * inverse document frequency
- For each content, calculate the vector of all tf*idf values
- Calculate cosine similarity of the vectors
Live demonstration
Abstracting activities into task descriptions

• Result
  – Activities with content that is associated to a chapter of the document
  – Activity consists of verb, object (content), target (person or system)
  – List of people that contributed to the document (in-)directly

• If enough activities of a type are found, they have to be aggregated
  – Results are generic task descriptions
    + content-specific parts based on headlines of the document or titles / keywords of the referenced content
    + role descriptions of the people (contributors) in relation to the document author (e.g. department head) or in general (e.g. KM expert)
From people to roles

- Possible relations on organizational level
  - Same department, same level
  - Same department, Superior/subordinate
  - Specific department (e.g. marketing)
  - Specific role / job description (e.g., project manager)

- Possible relations on the skill level
  - Same skill
  - Different skill
  - Specific skill set
People similarity / role abstraction

• Information from LDAP user directory are synchronized to neo4j
  – Result: graph with roles/job descriptions and org hierarchy derived from „manager“ relationships and department field

• Information from skill management system (Apache Shindig / neo4j)
  – Result: people – skill tag relationships

• Typical graph operations like shortest path between two people show relation in the organization hierarchy.

• Drawback: user directory and skill mgmt system have to be filled properly
Components and their roles

- **BPMN.io**: visualize & edit cases
- **Liferay**: store skills & activities
- **Nuxeo**: report doc events, associate case with doctype
- **Camunda BPM**: create case
- **Shindig + Mining**: find activities
- **ElasticSearch**: find relevant emails, indexing
- **Open LDAP**: sync org
- **Dovecot**: indexing
- **neo4j**:
Conclusion

• Combination of manual maturing steps (Web 2.0) and semi-automatic maturing (activity mining) seems promising

• Major challenges
  – Abstraction of concrete activities to general tasks
  – Abstraction of concrete users to formal roles

• (German) data protection laws have to be considered
  – Our solution does not present person-related data

• Evaluation with real life data is hard to perform

• CMMN editing capabilities for BPMN.io are on Camunda‘s roadmap
BPM + ACM + Web 2.0 + Data Analytics!