Fusepool:
Fusing and pooling information for product development

Dr. Michael Kaschesky – ksm1 [at] bfh.ch
Funding & consortium

[Logos of the organizations involved]

Web: www.fusepool.eu  Twitter: Fusepool
Overview

• **Introduction**: User-adaptive systems
• **Living Lab**: Rapid app development
• **Data processing**: Sourcing & interlinking
• **Machine learning**: Matching & optimizing
• **Sustainability**: Business plan & model
Introduction: User-adaptive systems

User-adaptive systems
Background

- SMEs lack resources to monitor and exploit
  - Technology intelligence for detecting and responding to opportunities and threats
    - Growth and complexity of patents and lawsuits
  - Consumer intelligence to detect opinions and needs of consumers for product development
  - Open innovation requiring cooperation (links between data, e.g. finding business partners)
- Focus: ML algorithms to improve matching

Web: www.fusepool.eu   Twitter: Fusepool
User-adaptive system

• **Focus**: monitor and learn specific needs and preferences of a user to align features, functionalities, and graphical interfaces

• **Adaptive**: machine learning from crowdsourcing (rather than ex-ante rule-based)

• **User-aligned prioritization**: more usable and customized interfaces, suggestions based on activity & im-/explicit feedback
User-adaptive matching

- **Main goal**: automated user-adaptive matching of users to funding opportunities
- **Key asset**: information provided by the user (behavior / crowdsourcing and uploads)
- **User data credo**: accuracy improves with quantity and quality of user data while variety (breadth) increases with number of users
- **Fusepool credo**: maximizing matching of content – not of advertisements

Web: www.fusepool.eu    Twitter: Fusepool
Berner Fachhochschule / Bern University of Applied Sciences
Fachbereich Wirtschaft / Bern Business School

Living Lab: Rapid app development
Rapid app development
Living lab & rapid app dev

- **Living lab**: Co-creation between producers and users of software
- **Rapid app dev**: Continuous prototyping and feedback from SMEs

Web: www.fusepool.eu   Twitter: Fusepool
Data processing: Sourcing & interlinking

Sourcing & interlinking
Data sourcing

• **Sources**: internal & external content from web harvesting & structured data sources (eg. research, patent databases, LOD)

• **Scope**: initial data corpus includes all explicitly in- and excluded sources in Google Custom Search API plus all other sources identified by Google (default)

• **Information gain value**: recommendations based on machine learning from feedback
Data handling

1. **Text feature extraction**: NLP methods for categorizing texts, entity recognition, etc.

2. **Shared metadata models**: mapping text features to existing/custom ontologies and generation of semantic triplets
   → high-level abstraction & persistence for reuse

   → **Lightweight storage**: mostly metadata only, text indexing and abstraction uses schema-free key-value (enabling actionable facets)

Web: www.fusepool.eu    Twitter: Fusepool
Data privacy

• **Goal:** data fusion from diverse sources without endangering user privacy
  – Maximize privacy by accounting for complex combinations of potentially identifying data
  – Minimize transformations of indirect data to maintain system accuracy and responsiveness

• **Metadata:** when a user uploads texts to be matched with other content, only the metadata descriptors are transmitted

Web: www.fusepool.eu  Twitter: Fusepool
Data interlinking

• **Contextualize**: terms are interlinked with same and similar terms across sources:
  – Enrich the extracted content with existing information available in the Internet
  – Interlink as much information as possible to increase the value of knowledge extraction
  – Use available public sector resources in Semantic Web and LOD format

• **Challenge**: ontology & taxonomy matching

Web: www.fusepool.eu    Twitter: Fusepool
Machine learning: Matching & optimizing

Matching & optimizing
Searching & finding

• **Key search-oriented features:**
  – Search through all content in the data pool
  – Faceted search (categories, metadata, entities)
  – Integration of Linked Open Data (LOD) results
  – Cross-lingual indexing and cross-referencing
  – “Did you mean?”-functionality in case of typos and auto-completion of search queries

• **User-adaptive:** indexing and integration based on user’s needs (e.g. user profiling)

Web: www.fusepool.eu   Twitter: Fusepool
Adaptation & refinement

- **Adaptive search**: results are aligned to user preferences based on analysis of user implicit and explicit feedback (learning to rank paradigm, e.g. Joachims & Radlinski)
- **Multi-task ranking**: good trade-off between user-independent search (high coverage but low precision) and fully customized systems
- **Query intent discovery**: structuring and interlinking an unstructured query input

Web: www.fusepool.eu    Twitter: Fusepool
Example: Query intent discovery

3

show journal articles on crowdsourcing of last 2 years

"journal articles" is recognized as source
"crowdsourcing" is recognized as topic
"last 2 years" is recognized as date range

FIND DOCUMENTS WHERE
SOURCE IS {ACM, DOAJ, ...}
TOPIC IS CROWDSOURCING
DATE RANGE IS 20090902 - 20110903
Correlating & matching

• **Search guided navigation**: semantic matching extracts contextual relationships to list related content
  – suggestions organized by categories
  – exposing facets within related content

• **Distributed rule and event model**: defines states, actions, and consequences (e.g. notifications, visualizations) for reasoning based on light-weight ontologies
Crowdsourcing & supervised automation

• **Relational learning**: related instances are used to reason about the focal instance
  – Relationality of content (links to other content, people, etc.) provide rich information
  – Similarities/dissimilarities to other content is established purely on relational properties

• **Tensor factorization**: matrix of terms with weights from annotated content is factored into a term matrix & content matrix/clusters
Sustainability: Business plan & model

Dr. Michael Kaschesky – ksm1 [at] bfh.ch
Business plan & model

- **Pricing model**: subscription model to generate income to maintain services
- **Licensing model**: Background IP of SME partners are used and compensated fair and reasonably
- **Customers**: SMEs and existing Living Labs and other open innovation system to support member SMEs

Web: www.fusepool.eu   Twitter: Fusepool
Thank you!

Web: www.fusepool.eu    Twitter: Fusepool

Dr. Michael Kaschesky – ksm1 [at] bfh.ch