Diary Generation from Personal Information Models to Support Contextual Remembering and Reminiscence

Christian Jilek, Heiko Maus, Sven Schwarz, Andreas Dengel

German Research Center for Artificial Intelligence (DFKI) Knowledge Management Department
1) Motivation / Background

- Vision
- PIMO & Semantic Desktop

2) Technical Realization

- User Interface (Client)
- Diary Generation (Server)

3) Early Evaluation

4) Conclusion & Outlook
Can you name **five things**

you were **concerned with the most**

for an **arbitrarily chosen period of your life**, 

e.g. September 2008 or spring 2003?
Vision

„a diary that writes itself“

on-demand diary generation from personal information models
PIMO represents the user’s mental model as vocabulary for applications w/o confronting users with the formal knowledge representation.
Ingredients of the Semantic Desktop Infrastructure

Dedicated PIMO Apps

Plug-in to (office) programs
Examples of Semantic Desktop Applications

Semantic Editor (SEED) [dedicated app]

FireTag for Mozilla Firefox [plug-in]
Vision (cont’d): Diversity within the Diary

<table>
<thead>
<tr>
<th>time</th>
<th>diary entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW 42</td>
<td>I was working at John Deere.</td>
</tr>
<tr>
<td>CW 43</td>
<td>I was working at John Deere. I attended the GSE lecture. I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 44</td>
<td>I attended the GSE lecture. I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 45</td>
<td>I was working at John Deere. I attended the GSE lecture. I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 46</td>
<td>I attended the GSE lecture. I attended the MMKI lecture. I attended the GRA lecture.</td>
</tr>
<tr>
<td>CW 47</td>
<td>I attended the GSE lecture. I attended the MMKI lecture. I attended the GRA lecture.</td>
</tr>
<tr>
<td>CW 48</td>
<td>I attended the GSE lecture. I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 49</td>
<td>I attended the GSE lecture. I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 50</td>
<td>I attended the GSE lecture. I attended the MMKI lecture. I attended the GRA lecture.</td>
</tr>
<tr>
<td>CW 51</td>
<td>I attended the GSE lecture.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>time</th>
<th>diary entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW 42-45</td>
<td>I was working at John Deere.</td>
</tr>
<tr>
<td>CW 43-50</td>
<td>I attended the MMKI lecture.</td>
</tr>
<tr>
<td>CW 43-51</td>
<td>I attended the GSE lecture.</td>
</tr>
<tr>
<td>CW 46-50</td>
<td>I attended the GRA lecture.</td>
</tr>
</tbody>
</table>
Vision (cont’d): Modern Look & Feel

screenshot by www.web-rater.com, 2014
PIMO Diary: User Interface
Structure of a Diary Entry

- date and headline
- associated photos
- most prominent things
- most prominent keywords
- most prominent annotations
Overall Context
## Overall Context

### 2012

- ADiWa
- Deutschland
- BMBF
- SmartOffice
- Köln
- ConTask
- ARIS
- Google Inc.
- Apple
- Innovation
- Semantic File Explorer
- Wissensmanagement
- Düsseldorf
- IEEE
- skipforward
- Infos zu gnwosis und elsevier Challenge für internen Newsletter
- ADiWa G4 Deliverable Evaluation
- CoPLANT

### 2014

- DFKI GmbH
- ForgetIT
- Sven Schwarz
- Kaiserslautern
- PIMO
- KM
- Andreas Dengel
- Ansgar Bernardi
- mobile
- Evaluation
- Deutschland
- EU
- ADiWa
- Cloud
- Ricoh
- Review
- App
- Personal Information Management
My Timeline of 2014

Select types to show (subtypes included):
- Agent
- Contract
- LifeSituation
- Location
- Project
- SocialEvent
- Topic
- Note

Show only my resources
Group Information MOdel (GIMO) Diary

Heiko’s initial entry

12.03.2014 - 31.03.2014
Related work DA Jilek identifizieren / Diplomarbeit Christian Jilek
Related work DA Jilek identifizieren
Vorarbeiten des DFKI mit Relevanz zu Diary
Jilek Konkretisierung Thema Diary
Jilek Kick-off
ChurchCherubiniOlive14: A large-scale study of daily information needs captured in situ
Diplomarbeit ab nächster Woche

Keywords:
jilek diary theme kick dfki related work woche daily large information diplomarbeit konkretisierung nächster study identifizieren vorarbeiten relevanz churchcherubiniolive14 scale captured situ even bahaa seed app alzayet sirkesk rahman pimostation pimomeeting michael kraus pimocal nops mohamed bedelidin

now including Christian’s shared data

12.03.2014 - 31.03.2014
Diplomarbeit Christian Jilek / diary
Related work DA Jilek identifizieren
Vorarbeiten des DFKI mit Relevanz zu Diary
Jilek Konkretisierung Thema Diary
Jilek Kick-off
Besprechung mit Heiko
Besprechung mit Heiko
Besprechung mit Heiko und Sven
Besprechung mit Sven
ChurchCherubiniOlive14: A large-scale study of daily information needs captured in situ
Related Work geprüft
Related Work
Besprechung mit Heiko
Besprechung mit Heiko
Diary Mockups
Besprechung mit Sven
Mehr Fragen
Fragen
JavaScript Tutorial
pim5 - smartoffice
Diplomarbeit ab nächster Woche

Keywords:
besprechung diary heiko even work related jilek fragen mockups app pimo dfki pimo5 seed tutorial diplomarbeit theme geprüft woche daily large information javascript kick 5 smartoffice buoyancy preservation konkretisierung nächster study identifizieren vorarbeiten relevanz churchcherubiniolive14 scale captured situ z.b 4
### Basic, Detail & Expert Settings

#### Start: 01.01.2014 06:00:00

**Time granularity:** year

**Discard least important:** false

**Data coverage:** 100%

---

**Detail Settings**

**Sorting:** blog

**Include shared:** false

**Entries desired:** 10

**Entries displayed:** 12

**Source items available:** 94

**Source items included:** 94

**Source items excluded:** 0

**Query duration:** 8 ms

- **Diversity:**
- **Rare Things:**
- **Rich Media Things:**

---

**Expert Settings**

- **Concepts in diary:** 5772
- **Concept index size:** 13770
- **Term index size:** 26501
- **Max. sim. found (weighted):** 0.90224
- **Bottler clusters in CC:** 0
- **Remove image things:** true
- **Use compact things:** true
- **Show cluster comp.:** false
- **Sim. threshold type:** amountOfMax
- **Sim. threshold value:** 0.25
- **Weight concepts:** 0.5

**Weight label terms:** 0.25

**Weight text terms:** 0.5

**Normalize sim. values:** true

**Max. weight related:** 0.75

**Centroid check (CC):** auto

**CC sim. thr. factor:** 2

**Find better clusters in CC:** false

**CC auto mode factor:** 1.5

**2nd Clust. Pass (2CP):** 1 rounded

**2CP ann. limit factor:** 0.75

**2CP incl. ann. things:** 10

**Ann. limit factor:** 0.75

**Rarity appl. factor:** 10

**Rarity bonus:** 1

**Rich media bonus:** 1

**Top elem. in context:** 20

**Label terms factor:** 3

**Keywords per entry:** 40

**Img. weight fav.:** 1

**Img. weight keep/del.:** 0.3

**Img. weight con. ann.:** 0.1

**Img. weight text ann.:** 0.1

**Img. weight quality:** 0.5

---

© DFKI - 2015
Technical Details at a Glance

• Similarity Calculation
  • using term vectors for headlines and text bodies
  • using concept vectors for concept annotations
  • Spreading Activation to find indirect annotations and
  • extend concept vectors

• Clustering of similar entries
  • also fosters higher diversity within the diary

• Importance evaluation

• Headline Generation
  next slides

• Text Summarization
Importance Evaluation

Importance evaluation of diary entries based on

- **annotation intensity**,  
- presence of things having a high potential of being a **memory landmark**,  
- **rarity** (idea: *rare persons/locations/etc. might be more memorable*),  
- associations with **rich media**

---

**09.08.2014 - 29.03.2014**

**Pizza**

composite entry consisting of 4 single entries:
- Pizzaessen [pimo:thing#Note]
- Pizza [pimo:thing#Topic]
- Pizzabacken [pimo:thing#Note]
- pizzabacken.jpg [pimo:thing#Image]

importance evaluation:
- (annotations | highPrioThings | rarity | richMedia) = (1.104 | 0.699 | 0.000 | 1.000)
- importance = 2.803

---

**27.06.2014**

**Spreading Activation**

single entry:
- Spreading Activation [pimo:thing#Note]

importance evaluation:
- (annotations | highPrioThings | rarity | richMedia) = (0.588 | 0.699 | 1.000 | 0.000)
- importance = 2.287
Headline Generation based on

- labels of **Annotations and Information Elements**
- intra-cluster importance evaluation (similar to previous slide)
- length of viewed time period
- possibly a split label

**Example:**

**ForgetIT / ForgetIT WS Luleå 2014**

**Keywords** as a summary of the entry

- weight( label terms ) > weight( text body terms )
Early Evaluation: Setting

- **PANiC:**
  - group of 4 participants (50%♀, 50%♂)
  - „PANiC“: acronym of their first names
  - Industrial Engineering students
  - in their last year before earning a master‘s degree
  - **4 months** access to our Semantic Desktop prototype
  - then testing **PIMO Diary** for **3 weeks**

- **Questionnaire:**
  - items 1-12: compact USE questionnaire (Lund, 2001)
  - items 13-20: specific questions concerning our app‘s core features
  - text field to express any kind of feedback or comments

- 7-point Likert scale, each item is phrased such that: 7: best value 1: worst value
Early Evaluation: Summary of Results

• **Overall ratings (80 items in total):**
  
  40x 7, 30x 6, 9x 5, 1x 4, 0x 3, 0x 2, 0x 1  → overall average rating: 6.36

• **Item 19:**
  The app allows for an appropriate and **satisfactory retrospection** on those parts of my life that are reflected by my PIMO.  → item‘s average rating: 6.75

• **Item 20:**
  The **overall context** provides a good impression, i.e. a quick overview, of those things (reflected by my PIMO) that concerned me the most in the chosen period.  → item‘s average rating: 6.75

• **Comments / Feedback of the PANiC group:**
  • „This program is very **innovative**, I havn‘t seen anything like this before.“
  • „I was surprised by its **intuitive handling** and the **quality of the results**.“
  • „Using the app was **fun**.“
  • „It‘s a nice add-on to the PIMO which **helps in keeping an overview**.“
Conclusion:
- app enabling *early contextualization*
- supporting and especially *easing personal retrospection*
- innovative app:
  - *self-writing diary* with *blog look & feel*
  - *diversity* to make reading more exciting
  - *zoom* in and out of *time periods*
  - *manually shift emphases* (experimental)
  - *overall context*
- promising results in a first user experience evaluation

Open issues / outlook on possible future work:
- text *summarization using natural language* (sentences)
- things having a *time span*
- more *social media* capabilities (diary sharing)
- algorithm and parameter *tuning*
- *topic lanes*
Future Work: Topic Lanes

- March
- April
- May
- June
- July
- August

- Master's thesis
- Database lecture
- Working at JD
- Sports club

Exam
Thank you for your attention!

😊

Any questions?

The work presented was partially funded by the European Commission in the context of the FP7 ICT project ForgetIT (under grant no: 600826).
References


• Sauermann et al., “Overview and outlook on the semantic desktop,” in Proc. of the 1st Workshop on The Semantic Desktop at ISWC, 2005.


• Eldesouky et al., “Seed, a natural language interface to knowledge bases,” in Proc. 17th Intl. Conf. on Human-Computer Interaction, Los Angeles, USA, 2015.


More References