Leveraging BPI with NLP and Organizational Semantics

Deniz Iren, Hajo Reijers
Vrije Universiteit Amsterdam

ICSSP, Paris, 2017
7.7.2017
INTRODUCTION

Continuous improvement is essential.

3 Challenges:
- Resolving requirement conflicts
- Impact analysis
- Actionable BPI planning

Automation with NLP and organizational semantics may help.
Large and complex organizations have hundreds of business processes.

Automation promises:

- Significant savings in effort
- Improvement in quality
HIGHLIGHTS OF P.I. PROCESS

Requirements elicitation

Requirements conflict resolution

Impact analysis

Define actionable BPI plan

BPI Plan
HIGHLIGHTS OF P.I. PROCESS

- Requirements are gathered from various sources,
- Different parties, different goals, unique perspectives.

- Requirements elicitation
- Requirements conflict resolution
- Impact analysis
- Define actionable BPI plan
- BPI Plan
HIGHLIGHTS OF P.I. PROCESS

- Requirements are gathered from various sources,
- Different parties, different goals, unique perspectives.

- Conflicts and duplicates may occur,
- Business analysts detect and resolve these conflicts among a large set of requirements.
HIGHLIGHTS OF P.I. PROCESS

- Identify the existing processes and systems that may be affected by these requirements,
- effort-intensive, manual, requires wide organizational knowledge.
HIGHLIGHTS OF P.I. PROCESS

- Define the scope of the BPI and an actionable plan
SOLUTION APPROACH

• Processes in the focus
• Process Models contain rich, linked data
• We extract and process this data
• Match requirements and various other artifacts to processes
**SOLUTION APPROACH**

**Process Model**
- Repository
- Realized requirements

**BPI Stages**
- Requirements elicitation
- Requirements conflict resolution
- Impact analysis
- Define actionable BPI plan

**Proposed approach**
- Map requirements to processes
- Identify potentially impacted organizational entities
- Visualize impact

**Impacted entities**

<table>
<thead>
<tr>
<th>Processes</th>
<th>Impact</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process A</td>
<td>✗</td>
<td>++</td>
</tr>
<tr>
<td>Process B</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Process C</td>
<td>✗</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Roles / Units</th>
<th>Impact</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role 1</td>
<td>✗</td>
<td>+++</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IT Systems</th>
<th>Impact</th>
<th>Magnitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>System 1</td>
<td></td>
<td>++</td>
</tr>
<tr>
<td>System 2</td>
<td>✗</td>
<td>+++++</td>
</tr>
</tbody>
</table>
SOLUTION APPROACH

BPI Stages

Requirements elicitation

Requirements conflict resolution

Impact analysis

Define actionable BPI plan

BPI Plan

Proposed approach

Process Model Repository

Realized requirements

Map requirements to processes

Identify potentially impacted organizational entities

Visualize impact

Organization impact

Process impact

VIEW1

VIEW2

VIEW3
# NLP FOR MATCHING

<table>
<thead>
<tr>
<th>ID</th>
<th>Requirement</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Customers shall be able to select multiple promotional packages for subscription</td>
<td>Marketing Dpt.</td>
</tr>
<tr>
<td>R2</td>
<td>StudentPack promotional package subscribers cannot subscribe to Senior package at the same time</td>
<td>Legal Dpt.</td>
</tr>
<tr>
<td>R3</td>
<td>Easy-CRM must display customers' contact information and the history of promotional packages subscribed, to the sales representative.</td>
<td>Customer Relations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Entity</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Easy-CRM</td>
<td>IT System</td>
</tr>
<tr>
<td>E2</td>
<td>TelePortal</td>
<td>IT System</td>
</tr>
<tr>
<td>E3</td>
<td>Online POS</td>
<td>IT System</td>
</tr>
<tr>
<td>E4</td>
<td>StudentPack</td>
<td>Product</td>
</tr>
<tr>
<td>E5</td>
<td>Sales Rep.</td>
<td>Role</td>
</tr>
</tbody>
</table>
EXAMPLE

Requirement statement:
“Customers shall be able to select multiple promotional packages for subscription.”

Subscription Process:
{{Subscription Process},
{TelePortal displays the list of promotional packages...},
{Select promotional package},
{TelePortal}, {Customer},...}
Requirement statement:

«Customers shall be able to select multiple promotional packages for subscription»

Objects/Subjects/Verbs:

Objects: {package, promotional packages, multiple promotional packages}

Subjects: {customers}

Verbs: {able, select, subscribe}
Requirement statement:
«StudentPack promotional package subscribers cannot subscribe to Senior package at the same time»

Bag-of-words:
Requirement statement: «StudentPack promotional package subscribers cannot subscribe to Senior package at the same time»

Sanitized Bag-of-words:

<table>
<thead>
<tr>
<th>Word</th>
<th>Tf</th>
<th>Idf</th>
<th>Tf x Idf</th>
</tr>
</thead>
<tbody>
<tr>
<td>StudentPack</td>
<td>1</td>
<td>1/25</td>
<td>1/25</td>
</tr>
<tr>
<td>Promotional</td>
<td>1</td>
<td>3/25</td>
<td>3/25</td>
</tr>
<tr>
<td>Package</td>
<td>2</td>
<td>4/25</td>
<td>8/25</td>
</tr>
<tr>
<td>Subscriber</td>
<td>1</td>
<td>1/25</td>
<td>1/25</td>
</tr>
<tr>
<td>Subscribe</td>
<td>1</td>
<td>3/25</td>
<td>3/25</td>
</tr>
<tr>
<td>Senior</td>
<td>1</td>
<td>1/25</td>
<td>1/25</td>
</tr>
<tr>
<td>Package</td>
<td>1</td>
<td>3/25</td>
<td>3/25</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
<td>1/25</td>
<td>1/25</td>
</tr>
</tbody>
</table>
Example

1. Represent in a canonical form
2. Identify lexical items
3. Lemmatize
4. Remove stop words
5. Calculate term frequency
6. Named entity recognition
7. Compare object-subject-verb sets
8. Calculate Cosine similarity
9. Compare semantic local context entity sets
10. Aggregate similarity measures

Subscription Process

Customer

Select promotional package

is eligible?

Subscribe
Pay

Task description: The TelePortal displays the list of promotional packages. The customer selects a package. The TelePortal checks the eligibility of the customer.

<table>
<thead>
<tr>
<th>ID</th>
<th>Entity</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Easy-CRM</td>
<td>IT System</td>
</tr>
<tr>
<td>E2</td>
<td>TelePortal</td>
<td>IT System</td>
</tr>
<tr>
<td>E3</td>
<td>Online POS</td>
<td>IT System</td>
</tr>
<tr>
<td>E4</td>
<td>StudentPack</td>
<td>Product</td>
</tr>
<tr>
<td>E5</td>
<td>Sales Representative</td>
<td>Actor</td>
</tr>
</tbody>
</table>
Represent in a canonical form
Identify lexical items
Lemmatize
Remove stop words
Calculate term frequency

Named entity recognition
Compare object-subject-verb sets
Calculate Cosine similarity
Compare semantic local context entity sets
Aggregate similarity measures

\[ JI'(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)} \]
Business Requirements

1. Represent in a canonical form
2. Identify lexical items
3. Lemmatize
4. Remove stop words
5. Calculate term frequency

Process Model Repository

6. Named entity recognition
7. Compare object-subject-verb sets

Semantic local context

8. Calculate Cosine similarity
9. Compare semantic local context entity sets
10. Aggregate similarity measures

EXAMPLE

| Word         | promotional | package | Senior | ...
|--------------|-------------|---------|--------|----------
| BoW-1        | 1/25        | 1/25    | 0      | ...      |
| BoW-2        | 0           | 0       | 3/25   | ...      |
| BoW-3        | 0           | 1/25    | 0      | ...      |
| BoW-4        | 1/25        | 0       | 0      | ...      |

\[ Sim_{Cos} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \]
EXAMPLE

1. Represent in a canonical form
2. Identify lexical items
3. Lemmatize
4. Remove stop words
5. Calculate term frequency
6. Named entity recognition
7. Compare object-subject-verb sets
8. Calculate Cosine similarity
9. Compare semantic local context entity sets
10. Aggregate similarity measures

\[ JI'(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)} \]
EXAMPLE

1. Represent in a canonical form
2. Identify lexical items
3. Lemmatize
4. Remove stop words
5. Calculate term frequency
6. Named entity recognition
7. Compare object-subject-verb sets
8. Calculate Cosine similarity
9. Compare semantic local context entity sets
10. Aggregate similarity measures

$$Sim = \frac{C_1 \cdot Sim_{Cos} + C_2 \cdot Sim_{LC} + C_3 \cdot Sim_{OS} + C_4 \cdot Sim_{V}}{C_1 + C_2 + C_3 + C_4}$$
EVALUATION

• Environment:
  – Real-life business setting (major telecom company)
  – BPMN repository (620 process models)
  – 8000 linked entities (local context)
  – 200 unique requirement statements

• Quantitative
  – Comparing the outputs of the automated approach with a gold standard manually created by business analysts.
  – Reporting Precision and Recall values

• Qualitative
  – Semi-structured interviews
QUANTITATIVE EVALUATION

• Gold Standard:
  – Matching between requirements and high-level processes

• Configuration:
  – 5 different configurations with different weight constants

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>LC</th>
<th>OS</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config-1</td>
<td>0.6</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Config-2</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-3</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-4</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
QUANTITATIVE EVALUATION

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>LC</th>
<th>OS</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Config-1</td>
<td>0.6</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Config-2</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-3</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-4</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Config-5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

![Graph showing precision-recall curves for different configurations with corresponding values in the table.](image-url)
**SEMI-STRUCTURED INTERVIEWS**

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Years of BPI Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior Analyst 1</td>
<td>17</td>
</tr>
<tr>
<td>Senior Analyst 2</td>
<td>5</td>
</tr>
<tr>
<td>Analyst</td>
<td>2</td>
</tr>
</tbody>
</table>

Interviews consisted of two stages
1) Walkthrough
2) Q/A session

- Description of the business analysts’ tasks in different BPI stages
- Effort requirements for these tasks
- Knowledge and expertise requirements to carry out these tasks
- The risks associated with these stages
- Opinion of analysts regarding the automated approach
SEMI-STRUCTURED INTERVIEWS

Conflict Resolution Phase

- Usually conducted by experienced analysts
- 30 – 45 person-minutes per requirement statement
- 20 minutes improvement is estimated per statement
- Increased consistency and quality, and reduced errors
Impact Analysis Phase

- Effort-intensive, time-consuming, and prone to errors
- Task was measured to take 120 person-hours
- An estimated 50% to 70% efficiency increase through automation
BPI Planning and Scope Determination

- For a small-medium size project this task takes approximately 80 person-hours
- 30% - 40% gain in time, and 20% decrease in errors are expected
- Enables less experienced personnel
SUMMARY

Punchline

• We achieved measurable improvement in time and quality

Future work

• Continuous requirements engineering for agile business settings?

Collaboration opportunities

• Contact
Questions / Contact Info

Deniz Iren
deniziren@gmail.com