

Generation of Recommendations for Process Templates Based on the Analysis of Change Histories

Sebastian Fröhlich

18.01.2016

Software Engineering for Business Information Systems (sebis)
Department of Informatics
Technische Universität München, Germany

www.matthes.in.tum.de

1. Introduction

1. Motivation
2. Objectives

2. Approach

1. Overview
2. Business Understanding
3. Data Understanding
4. Data Preparation
5. Modeling
6. Evaluation

3. Live DEMO

4. Discussion

SocioCortex:

» Sebis Public Website » Master's Thesis Sebastian Fröhlich

View Settings Files and Subpages Versions 0 Comments Watch

Master's Thesis Sebastian Fröhlich

Last modified by [Sebastian Fröhlich](#) Dec 16, 2015

[guided research](#) [michel](#) [masterthesis](#) [bpm](#) [case management](#)

Abstract Edit Wiki

Ad hoc processes have become a big role in modern organizations. Nevertheless, traditional workflow solutions have problems to support those processes. As a result, there is the trend moving away from those inflexible processes to collaborative solutions. Here contributors are able to design or customize existing steps for different activities, with their domain specific knowledge. Here the problem arises, that these flexible and maybe unstructured systems lead to a higher effort for recurrent processes, because every step has to be rethought and created repeatedly for similar cases. Additionally, unskilled users may be overstrained, because of the lack of guidance. Then, experienced users with their domain specific knowledge are needed to complete a process successfully. However, these

[<https://wwwmatthes.in.tum.de/pages/10mvmv60zxxk/Master-s-Thesis-Sebastian-Froehlich>]

Attributes of this Student Project Edit All

Title (de)	Generierung von Vorschlägen für Prozesstemplates basierend auf der Analyse von Änderungshistorien
Title (en)	Generation of Recommendations for Process Templates Based on the Analysis of Change Histories
Project	SocioCortex
Type	Master's Thesis
Status	completed
Student	Sebastian Fröhlich
Advisor	Felix Michel

- Collaborative information system developed by SEBIS chair
- Wiki with capability to store data and knowledge in semi-structured form

Attributes of this Student Project

Edit All

Task ?

Title (de)	Generierung von Vorschlägen für Prozesstemplates basierend auf der Analyse von Änderungshistorien
Title (en)	Generation of Recommendations for Process Templates Based on the Analysis of Change Histories
Project	SocioCortex
Type	Master's Thesis

Task ?

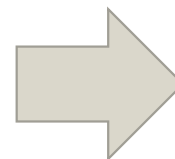
Status	completed
Student	Sebastian Fröhlich
Advisor	Felix Michel
Supervisor	Prof. Dr. Florian Matthes
Start Date	15.05.2015

Sebis Contributor Agreement signed on	12.05.2015
Checklist filled	Yes
Submission date	15.12.2015
Kick-off presentation slides	Initial_froehlich.pdf ⓘ
Final presentation slides	
Thesis PDF	MA_froehlich.pdf ⓘ
Copyright agreement - publication allowed	Yes
Copyright agreement - notification email	No
Copyright agreement - notification required	No <input type="button" value="Edit"/>
Copyright agreement required	Yes <input type="button" value="Edit"/>

- Creation of templates needs domain knowledge
- Unrealistic that one owner has experience in all sub-processes

Modelers need to:

- Ask domain experts or
- Analyse process data manually



Expenditure of time

Objective

Attributes of this Student Project Edit All

Task 2	
Title (de)	Generierung von Vorschlägen für Prozesstemplates basierend auf der Analyse von Änderungshistorien
Title (en)	Generation of Recommendations for Process Templates Based on the Analysis of Change Histories
Task 1	
Project	SocioCortex
Type	Master's Thesis
Status	completed
Student	Sebastian Fröhlich
Advisor	Felix Michel
Supervisor	Prof. Dr. Florian Matthes
Start Date	15.05.2015

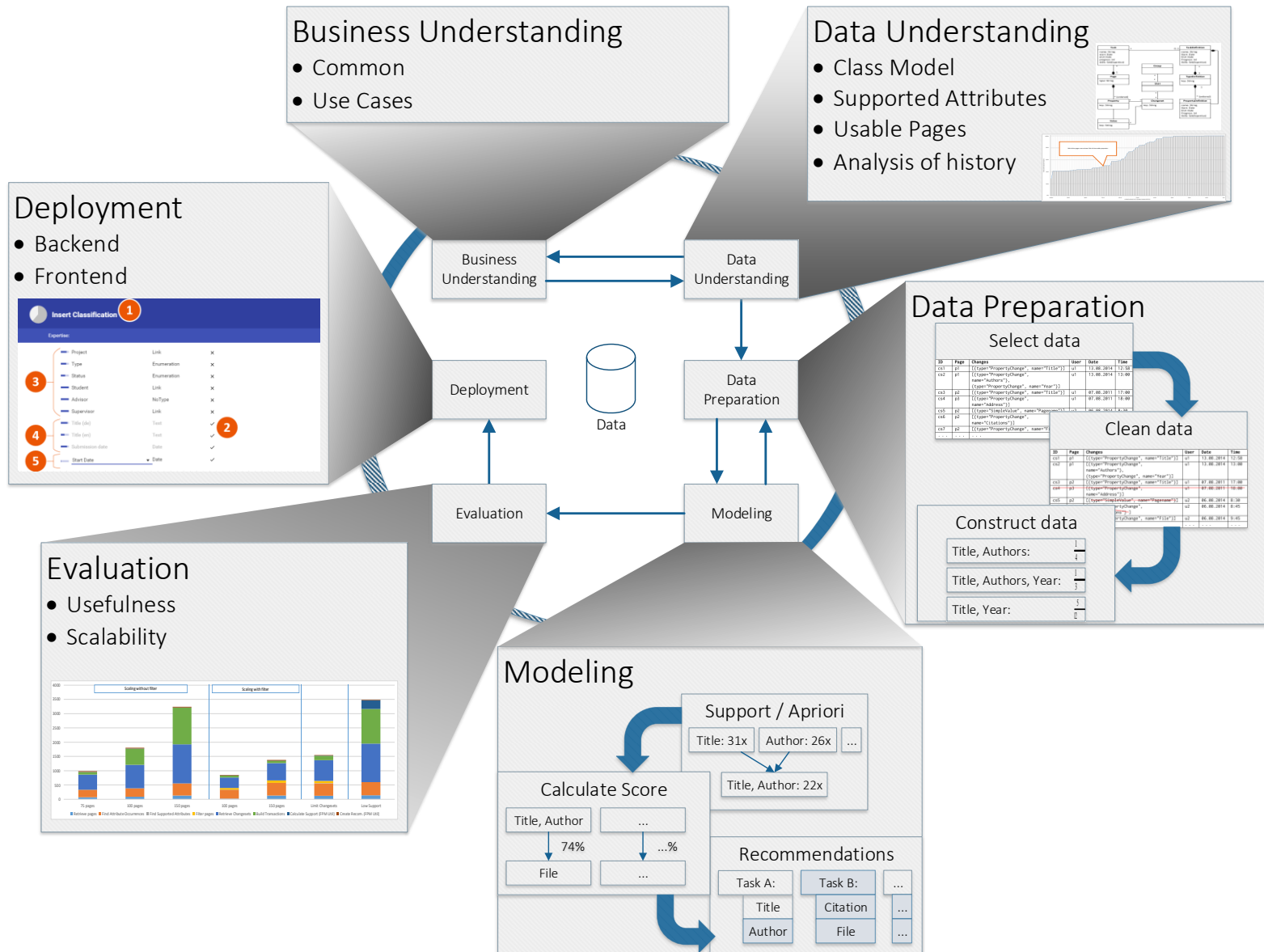
Task 4

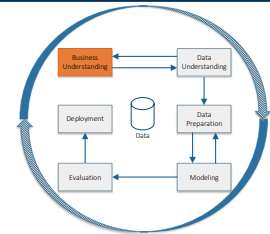
Sebis Contributor Agreement signed on	12.05.2015
Checklist filled	Yes
Submission date	15.12.2015
Kick-off presentation slides	Initial_froehlich.pdf ⓘ
Final presentation slides	
Thesis PDF	MA_froehlich.pdf ⓘ

Task 3

Aline Schmidt-Nicic	Edited	Sebis Contributor Agreement signed on
Aline Schmidt-Nicic	Edited	Copyright agreement - publication allo
Aline Schmidt-Nicic	Edited	Copyright agreement required ⓘ
Sebastian Fröhlich	Edited	Title (de) ⓘ
Sebastian Fröhlich	Edited	Title (en) ⓘ
Sebastian Fröhlich	Edited	Description ⓘ
Felix Michel	Edited	Advisor ⓘ
Felix Michel	Edited	Status ⓘ
Felix Michel	Edited	Submission date ⓘ
Felix Michel	Edited	Start Date ⓘ

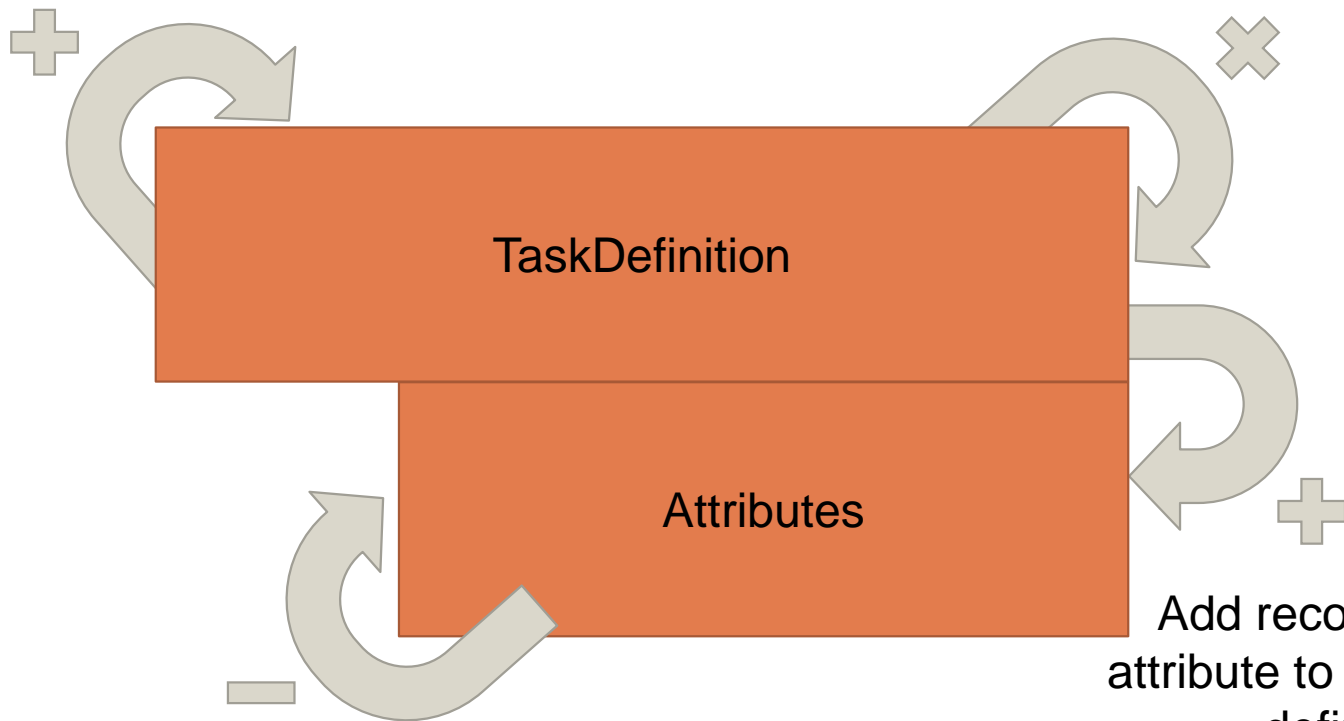
Support users to model tasks with attributes in context of socio cortex





Create new task definition with attributes

Remove existing task definition

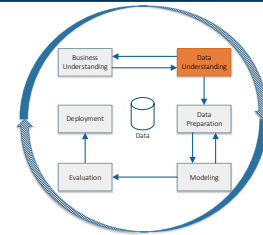


Add recommended attribute to existing task definition

Remove attribute from task definition

Sample usage for type **Publication**:

Usage of attribute	Attribute	Has AttributeDefinition
99.2 %	Year	✓
98.3 %	Title	✓
95.8 %	Key	✓
95.0 %	Citation	✓
89.9 %	Authors	✓
69.7 %	Address	✓
42.0 %	Type of publication	✓
39.5 %	Published in	X
17.6 %	Research project	✓
10.9 %	File	✓
24.4 %	Acronym	X
4.2 %	Type	X
2.5 %	Team members	X
...



Problem:

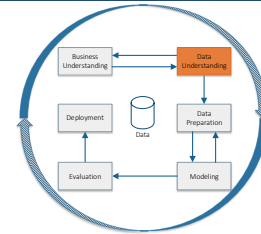
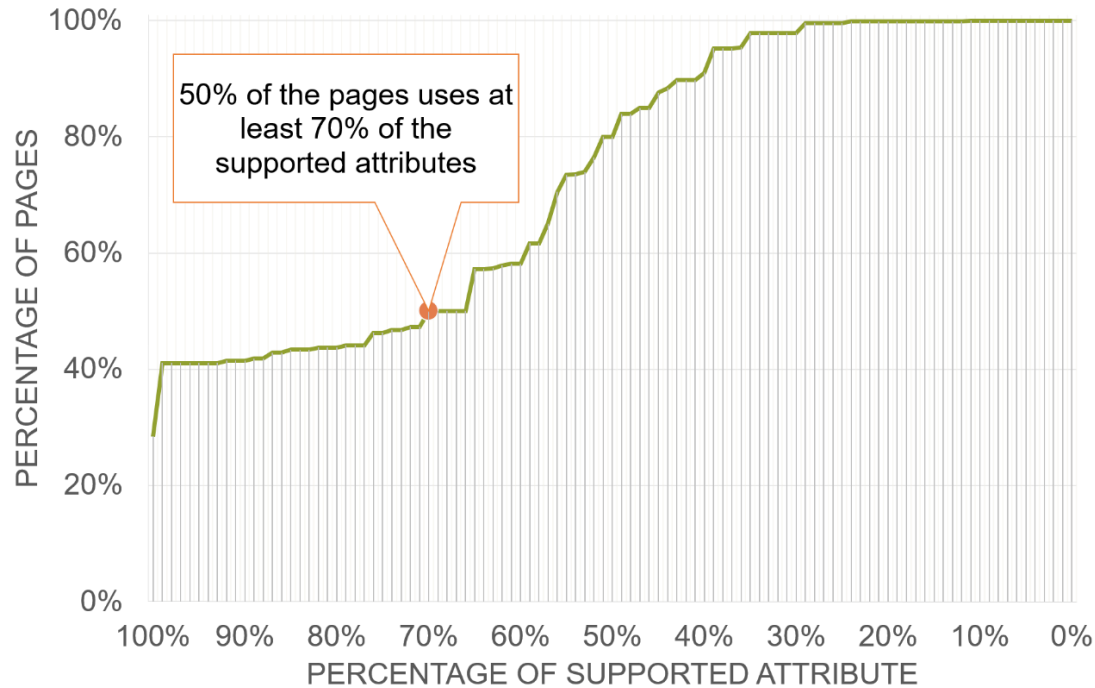
Which attributes should be used for creating recommendations?

Perception:

- Many free attributes exist on a lot of pages
- Some pages have no predefined AttributeDefinitions
- Types exist, where almost all pages have same free attributes

Result:

Free attributes need to be considered for recommendation



Problem:

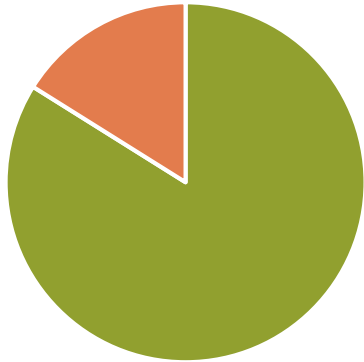
Incomplete pages may falsify result

Perception:

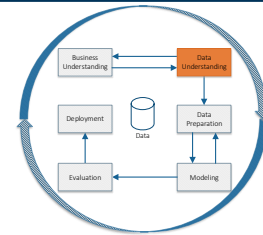
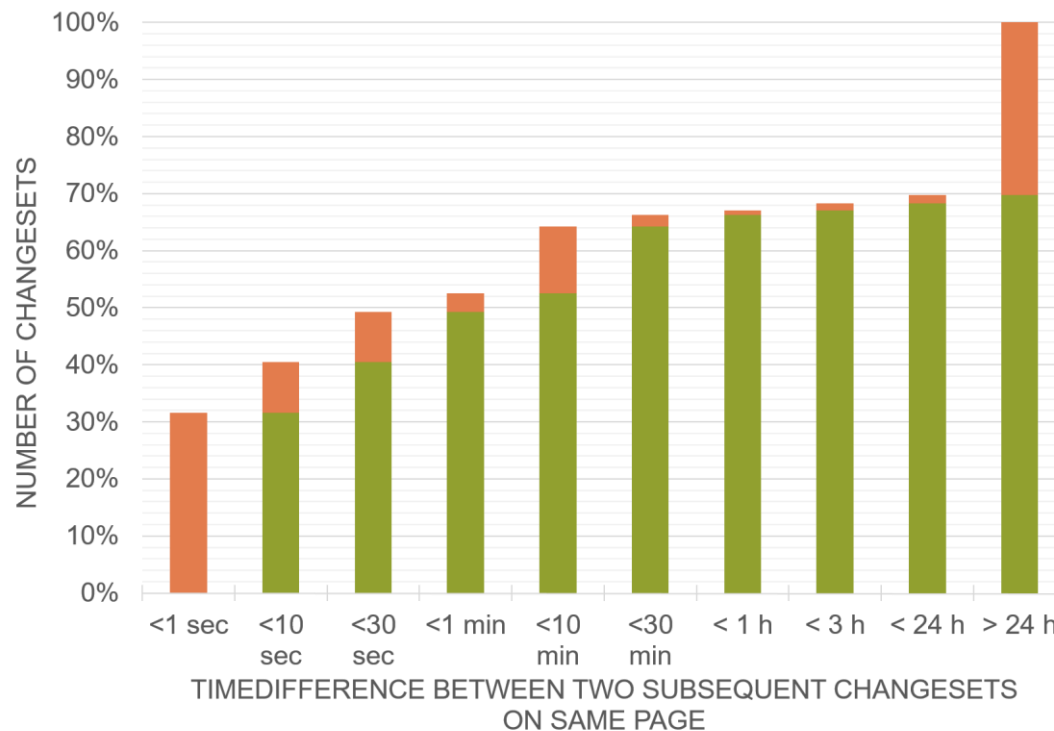
- Many pages are incomplete
- 50% of the pages uses at least 70% of the supported attributes

Result:

Use subset of pages (without incomplete ones)



- Changesets with single Change
- Changesets with multiple changes



Background:

Several methods to edit attributes

- Batch job
- Edit mode
- Single change

Problem:

Over 80% of changesets contain only 1 change

Perception:

- Almost 1/3 of all changes within 1 second
- Most changes within 30 minutes
- 2/3 of all changes done within 1 day

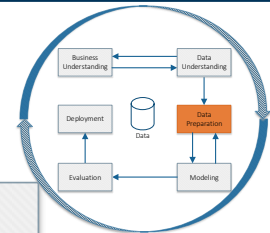
Result:

Join changesets

1. Retrieve pages for type
2. Find attribute occurrences
3. Find supported attributes
4. Retrieve Changesets

Filter changes which:

- refer to incomplete pages
- do not modify an attribute
- do not apply to supported attributes

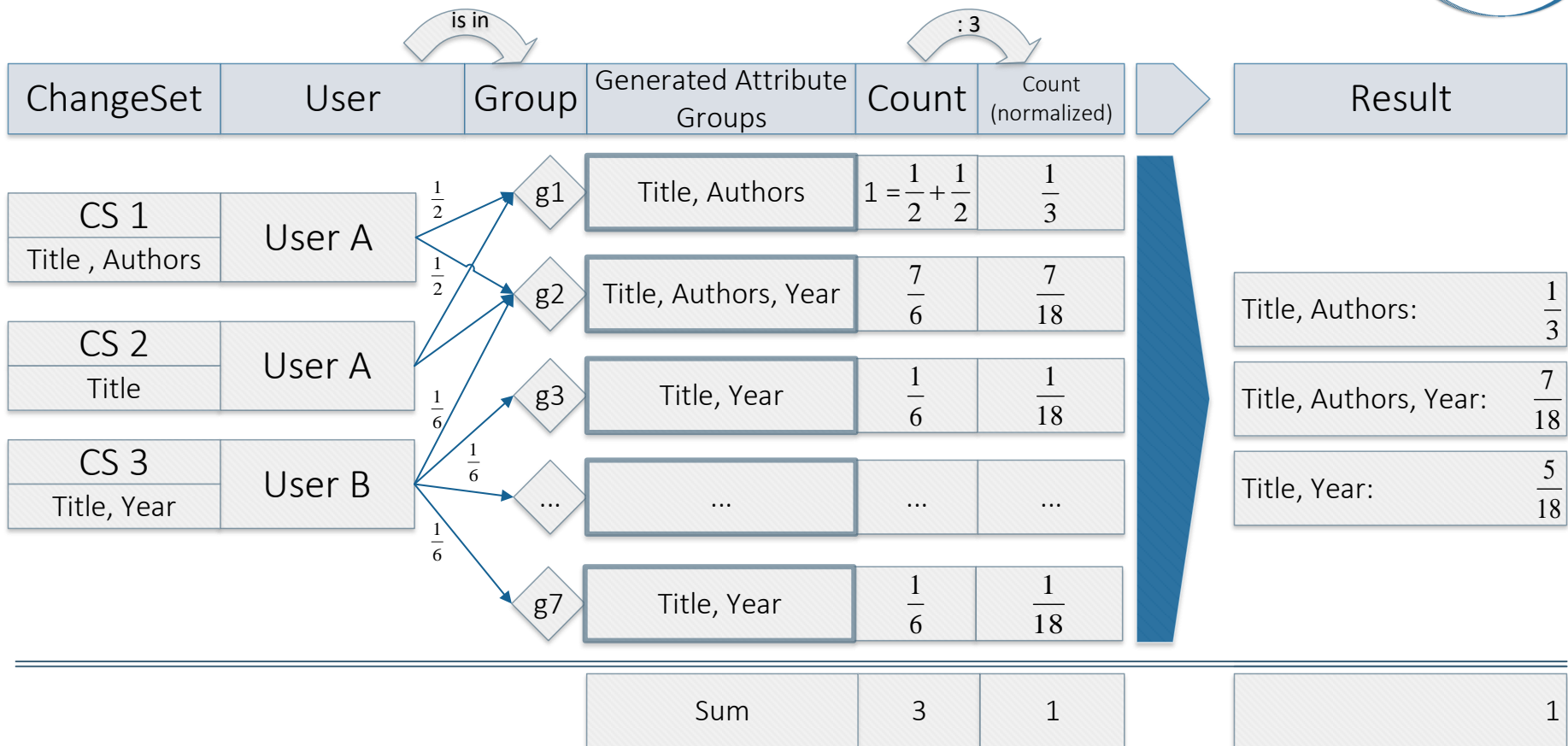
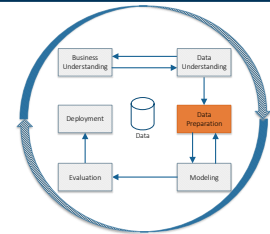


Select data					
ID	Page	Changes	User	Date	Time
cs1	p1	[{type="PropertyChange", name="Title"}]	u1	13.08.2014	12:58
cs2	p1	[{type="PropertyChange", name="Authors"}, {type="PropertyChange", name="Year"}]	u1	13.08.2014	13:00
cs3	p2	[{type="PropertyChange", name="Title"}]	u1	07.08.2011	17:00
cs4	p3	[{type="PropertyChange", name="Address"}]	u1	07.08.2011	18:00
cs5	p2	[{type="SimpleValue", name="PageName"}]	u2	06.08.2014	8:30

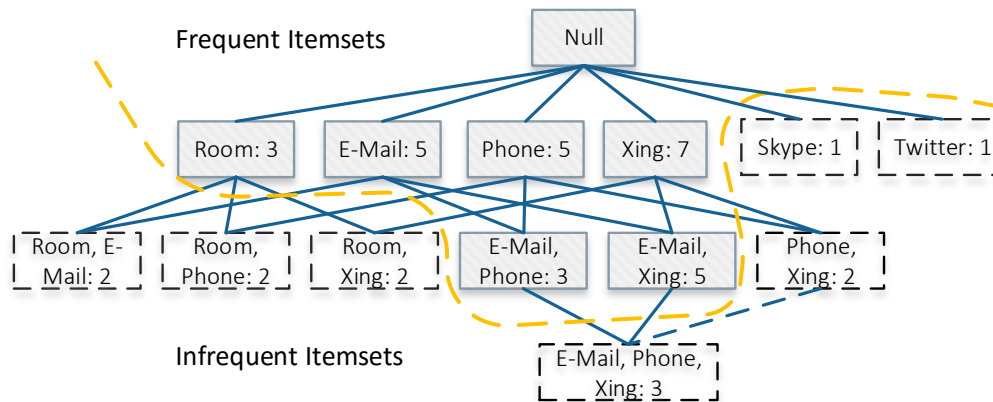
Clean data					
ID	Page	Changes	User	Date	Time
cs1	p1	[{type=AttributeChange, name=Title}]	u1	13.08.2014	12:58
cs2	p1	[{type=AttributeChange, name=Authors}, {type=AttributeChange, name=Year}]	u1	13.08.2014	13:00
cs3	p2	[{type=AttributeChange, name=Title}]	u1	07.08.2011	17:00
cs4	p3	[{type=AttributeChange, name=Address}]	u1	07.08.2011	18:00
cs5	p2	[{type=SimpleValue, name=PageName}]	u2	06.08.2014	8:30
cs6	p2	[{type=AttributeChange, name=Citations}]	u2	06.08.2014	8:45
cs7	p2	[{type=AttributeChange, name=File}]	u2	06.08.2014	9:45
...

Construct data	
Title, Authors:	$\frac{1}{4}$
Title, Authors, Year:	$\frac{1}{3}$
Title, Year:	$\frac{5}{12}$

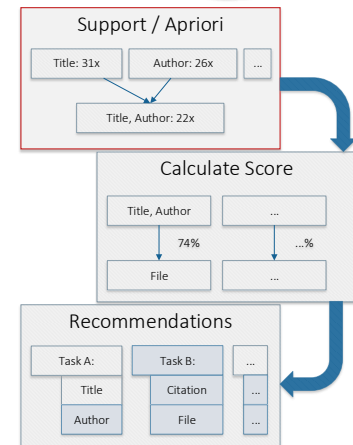
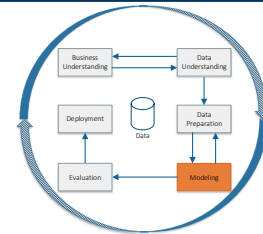
Join Changesets by user groups within certain time intervall



1. Create all frequent 1-itemsets
2. Join all (k-1) itemsets which differ in 1 item
3. Test if all subsets are frequent
4. Calculate support
5. Perform it again (until no new (k+1) itemset is found)



Example „Team Member“



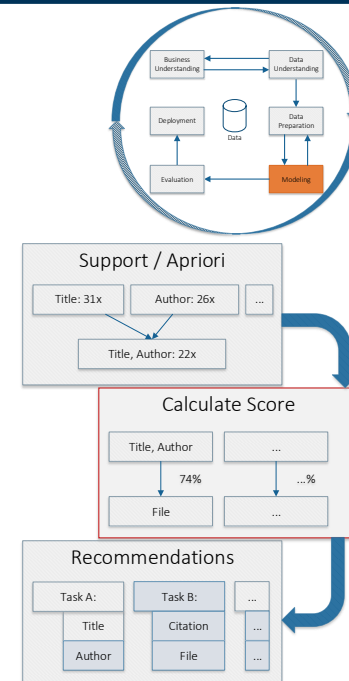
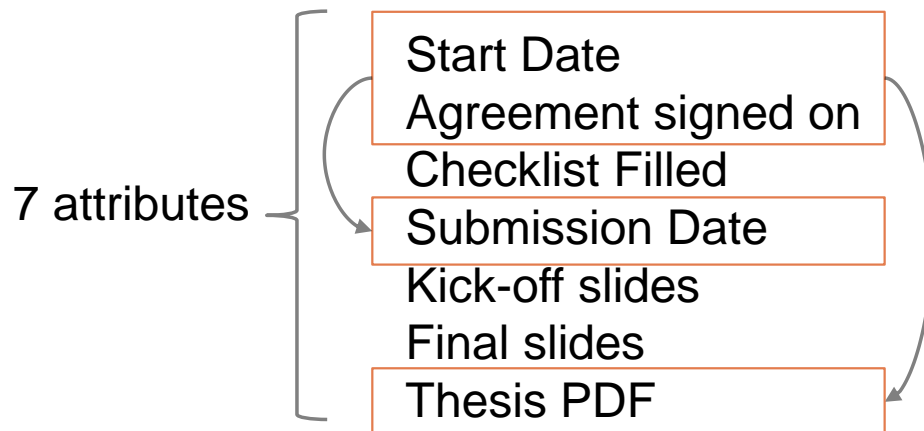
Modeling – Calculate Score

$minDist :=$ minimal distance between tested attribute and the attribute order of page

$$adjustment = \begin{cases} 0, & \text{if } minDist = MAX_VALUE \\ maxFactor - \left(\frac{2 * maxFactor * (minDist - 1)}{maxDist - 1} \right), & \text{otherwise} \end{cases}$$

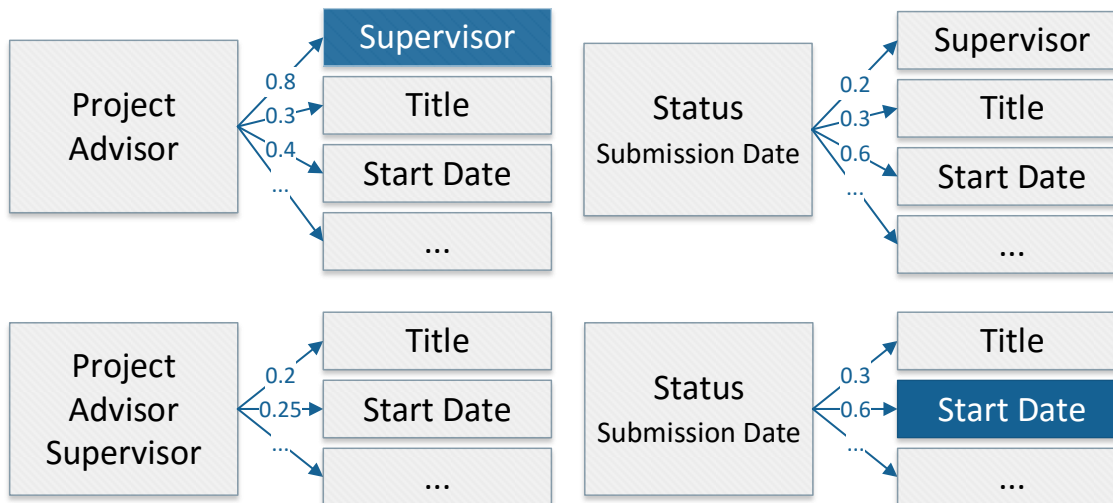
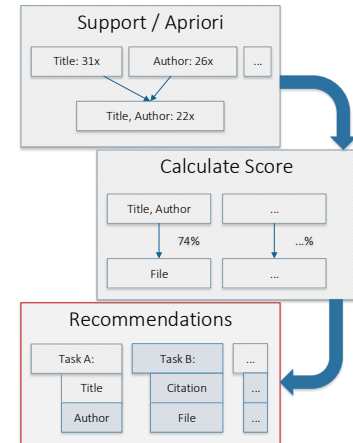
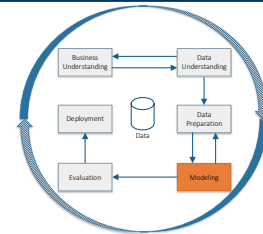
$$score = \max(\underbrace{\min(avgConf * (1 + adjustment), 1)}_{score \in [0; 1]}, 0)$$

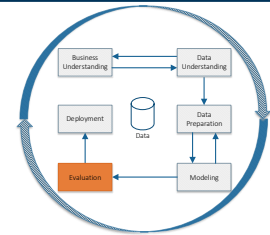
$adjustment \in [-maxFactor; maxFactor]$



Modeling – Create Recommendations

1. Find best attribute / existing TaskDefinition combination
2. „Add“ best matching attribute as recommendation
3. Repeat this until no combination reaches a specified minScore / minSupport
4. Pick up remaining attributes and build new recommendation
5. Repeat step 2,3 for it
6. Repeat 4,5 until no attribute is left





In task	Score	Attribute
Recommended Task		
X	0.4	Project
X	1.0	Advisor
X	1.0	Supervisor
Recommended Task		
X	1.0	Title(de)
X	1.0	Title (en)
X	0.8	Student
X	0.7	Start Date
Recommended Task		
X	0.5	Submission date
X	0.3	Thesis PDF

In task	Score	Attribute
Setup assignment		
✓	0.3	Project
✓	0.6	Advisor
X	0.8	Supervisor
Enter organizational aspects		
✓	0.6	Status
✓	0.7	Student
✓	0.4	Submission date
X	0.6	Start date
Enter title		
✓	0.5	Title (de)
✓	0.3	Title (en)
✓	0.1	Thesis PDF

- Dashboards
 - Activity Stream
 - Completeness
 - Statistics
 - Inconsistencies
 - Create new Type
- Types
 - Alumni
 - Alumni2
 - Anleitung
 - Article
 - Bachelorthesis
 - Book
 - Course
 - EA Vistype
- Workspace Actions
 - Import/Export
 - Settings

GENERAL TASKS



Setup Assignment

Expertise:

Project	Link	x
Advisor	NoType	x
Supervisor	Link	x

Route



Enter new taskname



Enter new taskname

Title (de)	Text	x
Title (en)	Text	x
Student	Link	x
Start Date	Date	x

Live DEMO

Discussion



Sebastian Fröhlich



Technische Universität München
Department of Informatics
Chair of Software Engineering for
Business Information Systems

Boltzmannstraße 3
85748 Garching bei München

Tel +49.89.289.17129
Fax +49.89.289.17136

www.matthes.in.tum.de

Sample usage for type **Article**:

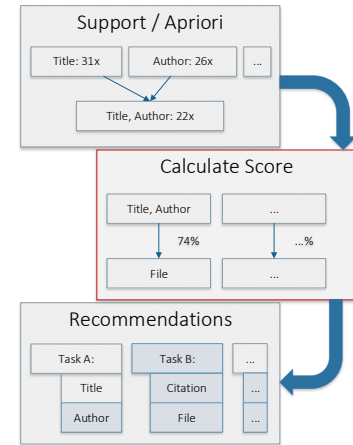
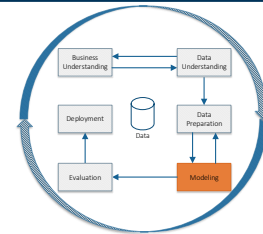
Usage of attribute	Attribute	Has AttributeDefinition
100.0 %	Key	X
100.0 %	Year	X
98.9 %	Citation	X
98.9 %	Title	X
96.7 %	Authors	X
90.0 %	Address	X
66.7 %	Published in	X
40.0 %	Research project	X
7.8 %	File	X
...

Modeling – Calculate Score

$minDist :=$ minimal distance between tested attribute and the attribute order of page

$$adjustment = \begin{cases} 0, & \text{if } minDist = MAX_VALUE \\ maxFactor - \left(\frac{2 * maxFactor * (minDist - 1)}{maxDist - 1} \right), & \text{otherwise} \end{cases}$$

$$score = \max(\min(avgConf * (1 + adjustment), 1), 0)$$



7 attributes

- Start Date
- Agreement signed on
- Checklist Filled
- Submission Date
- ...

$$maxFactor = 0.2, minDistance = 2$$

$$adjustment = 0.2 - \left(\frac{2 * 0.2 * (2 - 1)}{(7 - 1)} \right)$$

$$adjustment = 0.2 - 0.67$$

$$adjustment = 0.13$$

$$Confidence([Start Date, Agreement signed on] \Rightarrow [Submission Date]) = 0.6$$

$$Confidence([Submission Date] \Rightarrow [Start Date, Agreement signed on]) = 0.4$$

$$score = \max(\min(0.5 * (1 + 0.13), 1), 0) = 0.57$$

I

- Completeness
- Statistics
- Inconsistencies
- Create new Type
- Types
 - Person
 - Project
 - Publication
 - Research Project
 - Seminar
 - Staff
 - Student Project**
 - Team Member
- Workspace Actions
 - Import/Export
 - Settings

II

Insert Classification

Expertise:

Project	Link	x
Type	Enumeration	x
Status	Enumeration	x
Student	Link	x
Advisor	NoType	x
Supervisor	Link	x
Title (de)	Text	✓
Title (en)	Text	✓
Submission date	Date	✓
Add other attribute		

Use case 4

Use case 3

Use case 2

III

Enter new taskname

IV

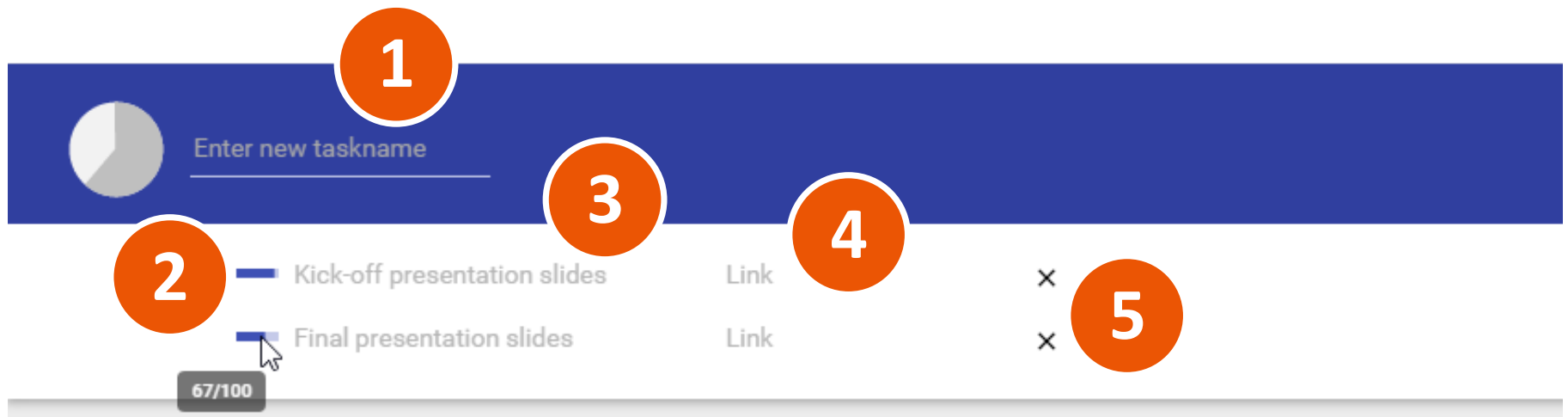
Enter new taskname

Kick-off presentation slides	Link	x
Final presentation slides	Link	x

Enter new taskname

Checklist filled	Enumeration	x
Sebis Contributor Agreement signed on	Date	x

Use case 1





Insert Classification

1

Expertise:

2	<input type="checkbox"/> Project	Link	×
	<input type="checkbox"/> Type	Enumeration	×
	<input type="checkbox"/> Status	Enumeration	×
	<input type="checkbox"/> Student	Link	×
	<input type="checkbox"/> Advisor	NoType	×
	<input type="checkbox"/> Supervisor	Link	×
3	<input type="checkbox"/> Title (de)	Text	✓
	<input type="checkbox"/> Title (en)	Text	✓
	<input type="checkbox"/> Submission date	Date	✓
4	<input type="checkbox"/> Start Date	▼ Date	✓

5

