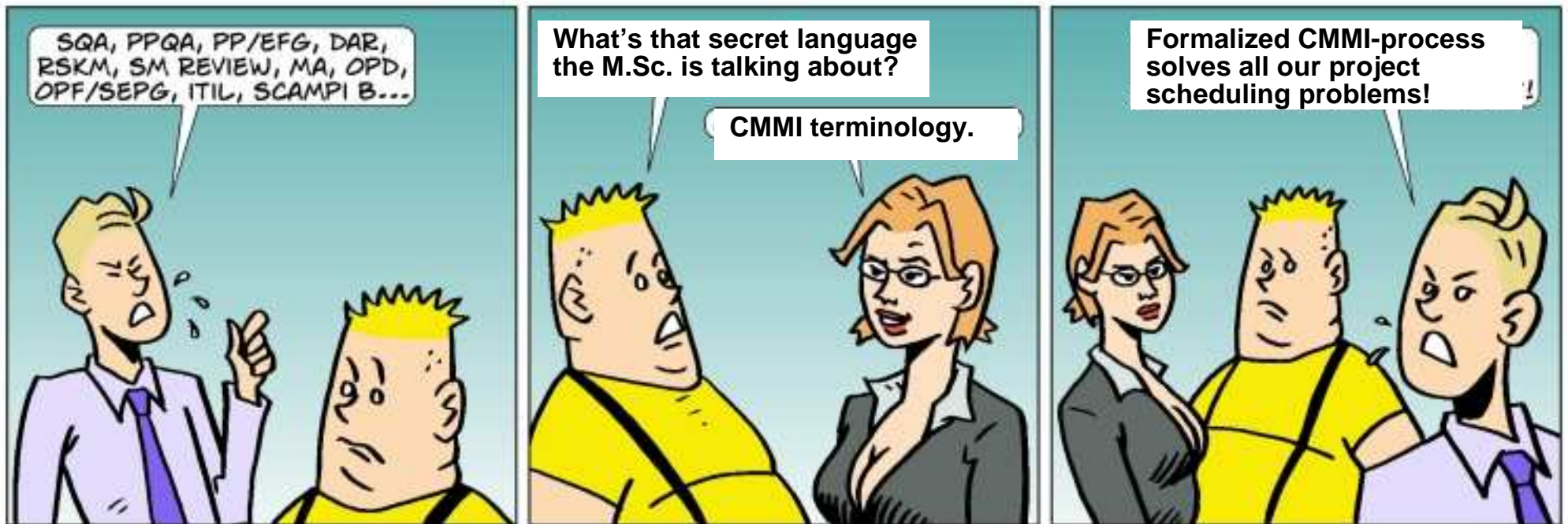




Workshop on Text Data Mining and Management (TDMM)
April 15, 2007, Istanbul, Turkey

ProcMiner: Advancing Process Analysis and Management

Miika Nurminen
Anne Honkaranta
Tommi Kärkkäinen
*Faculty of Information Technology
University of Jyväskylä, Finland*



Original text: OSSU MÄNTYLÄHTI Art: JUKKA PIIRA

Background

- Organizations utilize process models for various purposes
 - Business process re-engineering (reorganizing & automating work)
 - Process-aware systems (content & workflow management, ERP, SOA...)
 - Establishing a quality system (ISO 9001, EFQM, CMMI, ITIL...)
- Formality and specificity of process models varies
 - Visual graphs (Visio drawings, flowcharts, "swimlanes", UML)
 - Informal text descriptions (e.g. textual use cases)
 - Semistructured models (ProcML, QPR)
 - Formal, executable models (BPEL, XPDL)
- Challenges in process management
 - The more expressive process model, the more complex modeling process
 - Imprecise & ambiguous models, varying conventions & terminology
 - Incorporating process models to operational work
 - Maintaining models as processes change (and vice versa)

Text Mining for Process Management

- Process mining has mainly been applied to reverse the process of constructing the workflow model on design phase (e.g. workflow logs are used to construct a process specification).
- Novel information can also be discovered by applying text mining to collections of process models on design phase
 - Grouping processes by clustering, model reuse, enhanced search
 - Discovering “hot spot” actors or documents from process models
 - Optimizing process structure with structured text mining
- A new categorization for process mining is required
 - Following the popular web mining categorization (Madria *et al*, 1999), we distinguish *process content*, *structure* and *usage* mining.
 - Traditional process mining can be classified as process usage mining
 - Process content and structure mining produces patterns about process models, not the models themselves

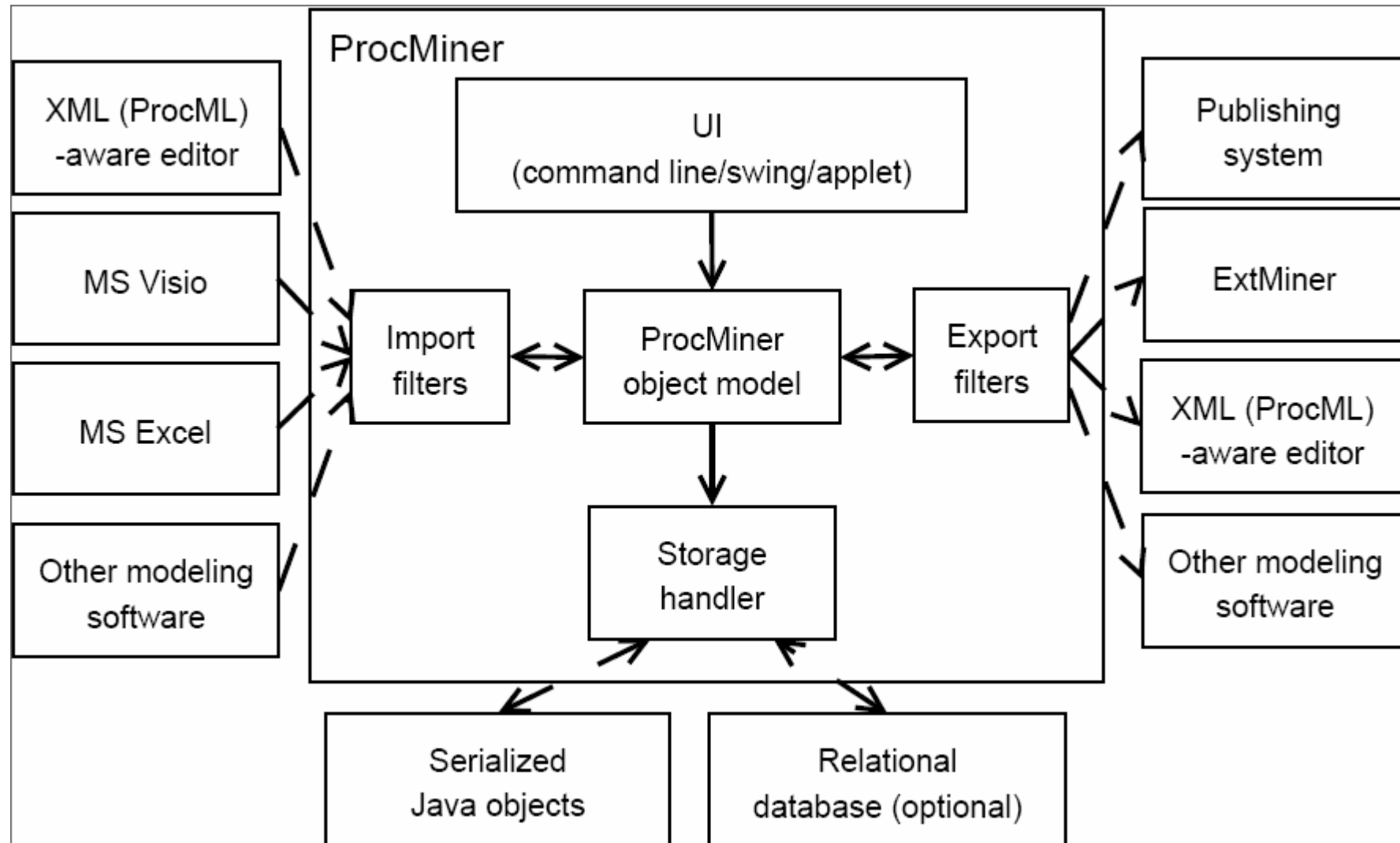
Related work

- Business Process Management, Process Mining (van der Aalst *et al*), workflow usage mining, patterns
- MIT Process Handbook (Malone *et al*, 2003) informal, yet structured approach for process modeling
- Workflow modeling (Sharp & McDermott, 2001) *swimlane-oriented process modeling techniques*
- (Cockburn, 2000), process (or use case) models with multiple abstraction levels
- (Ellmer & Merkl., 1996) example of content-based (software) process model clustering
- ExtMiner (Nurminen *et al*, 2005) a platform for searching & clustering structured documents

ProcMiner

- ProcMiner facilitates gathering process model information and producing novel combinations of information residing in the contents of the process models
 - XML-based process markup language based on an intermediate object model that is convertible to many process representations.
 - Versatile process retrieval and publishing functionality.
 - Support to process mining (content-based document clustering) by using ExtMiner, a platform for structured document retrieval and text mining.
- Integrates features previously implemented in separate systems
 - e.g. BPM, text mining, structured document clustering, multichannel publishing, information retrieval
- ProcMiner was used in the process mining, modeling and development initiative in the Faculty of Information Technology, University of Jyväskylä

ProcMiner architecture



ProcMiner Architecture (decomposed)

- 3 layers: UI, Process model logic and Data storage.
- Process model can be serialized using standard Java object serialization mechanism, or optionally to a relational database.
- Process logic includes a core object model that can be interfaced with import- and export filters for additional data formats, external applications and functionality (e.g. publishing with process portal, process model clustering with ExtMiner).
- Can be used with a command-line interface, Swing-based desktop application or an applet-enhanced web portal.
- Implemented with Java and PHP, published as open source. Third-party open source components (eg. GraphViz, LaTeX) are utilized.

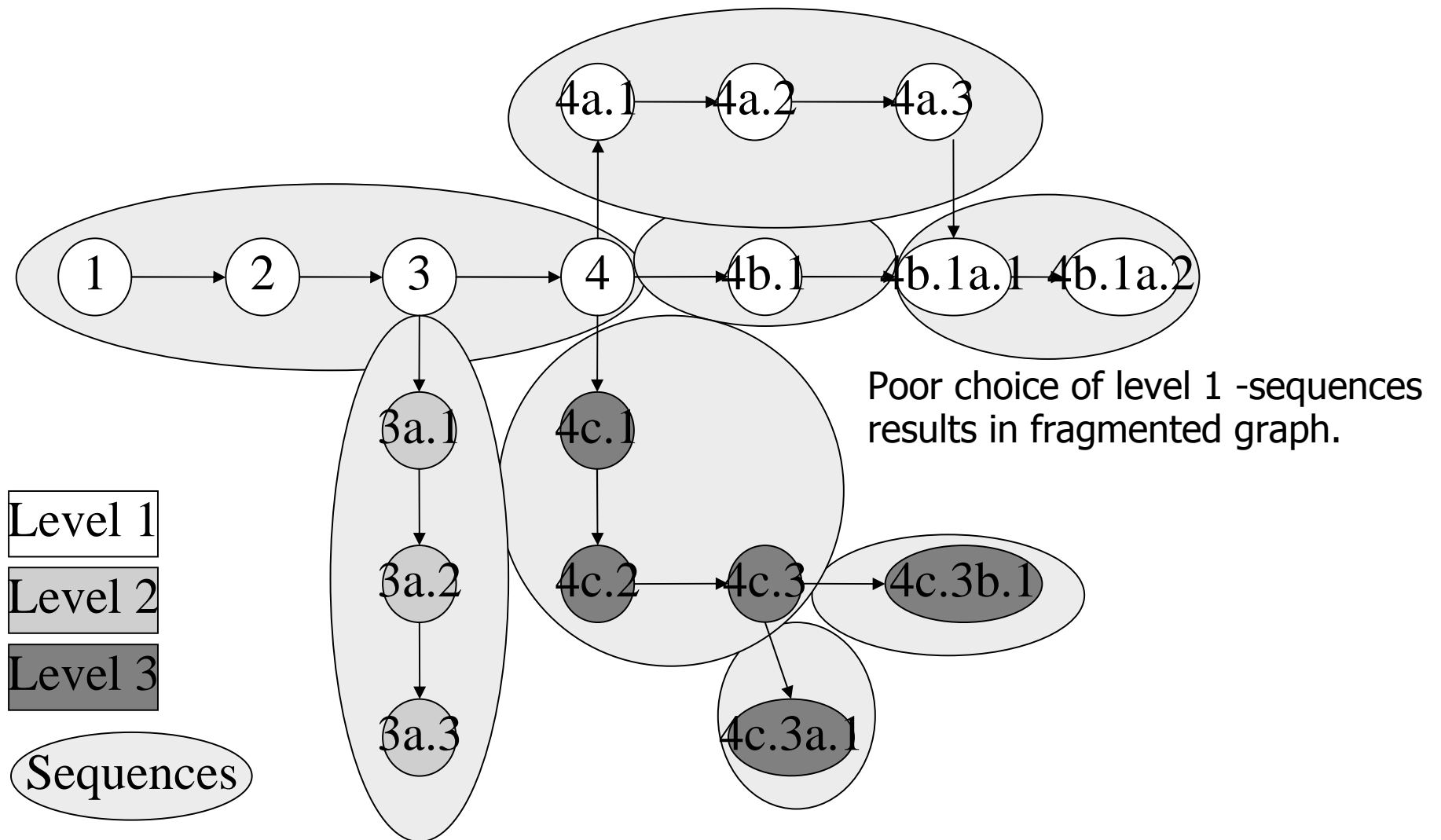
ProcMiner Object Model

- ProcMiner object model works as an intermediate format facilitating conversions between multiple modeling languages.
- Adaptable for different semiformal process models (i.e. structured models without formal semantics – cannot be executed, but are understandable and analyzable).
- Separation of process and process instance. Process is an abstract specification of the general characteristics related to a process. Process instance in an organization-specific model with additional metadata and a workflow graph.
- Process (instance) model is a multilevel graph, where each level adds more elements or overrides elements in the upper level. Subprocesses and links between process instances are also possible.
- Roles, documents and systems are modeled as trees or lists.

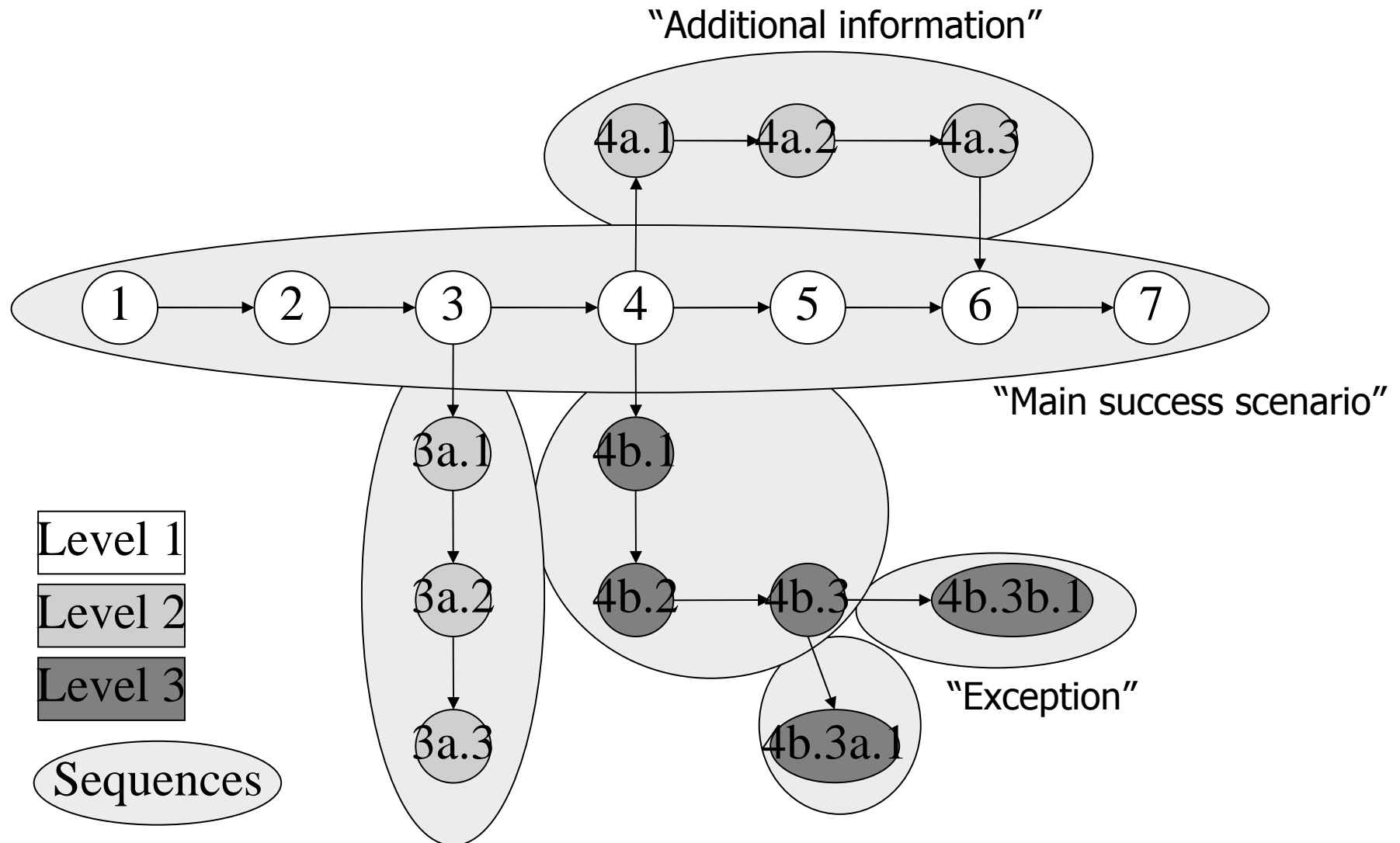
ProcML Modeling Language

- ProcMiner uses XML-based process modeling language ProcML that works as a human-readable format for object model.
- The language is designed for ease of expressivity for input of multilevel graph data without the need to use graphical tools. The graph is partitioned to both abstraction levels and *sequences*.
- Other process modeling languages (e.g. BPEL or XPDL) were considered to be too complex (and inadequate to express the new modeling concepts) for end-user driven modeling.
- Contrary to BPEL, ProcML is *not* designed to be executable. This simplifies the modeling, since many processes do not have to (nor even can be) automated.
- Despite the lack of formal semantics, ProcML models are structured and thus can be easily searched and maintained.

ProcML: Graph Partitioning



ProcML: Graph Partitioning (fixed)



Retrieval and Publishing

- Processes can be retrieved using full-text or metadata field - search, as well as browsing by document, role or information system lists that show all the processes where the given modeling entity is located.
- Both process metadata and graphical information is retrievable from the same object model. There is no need to maintain separate model and metadata documents.
- Publishing system produces a HTML-based "process portal" that contains a search engine, process descriptions and process-, document-, role-, and information systems trees or lists. Process descriptions contain both textual and graphical representation with automatic layout generated by Graphviz.
- For printing, a PDF-based "handbook" is generated using XSL Transformations and LaTeX.

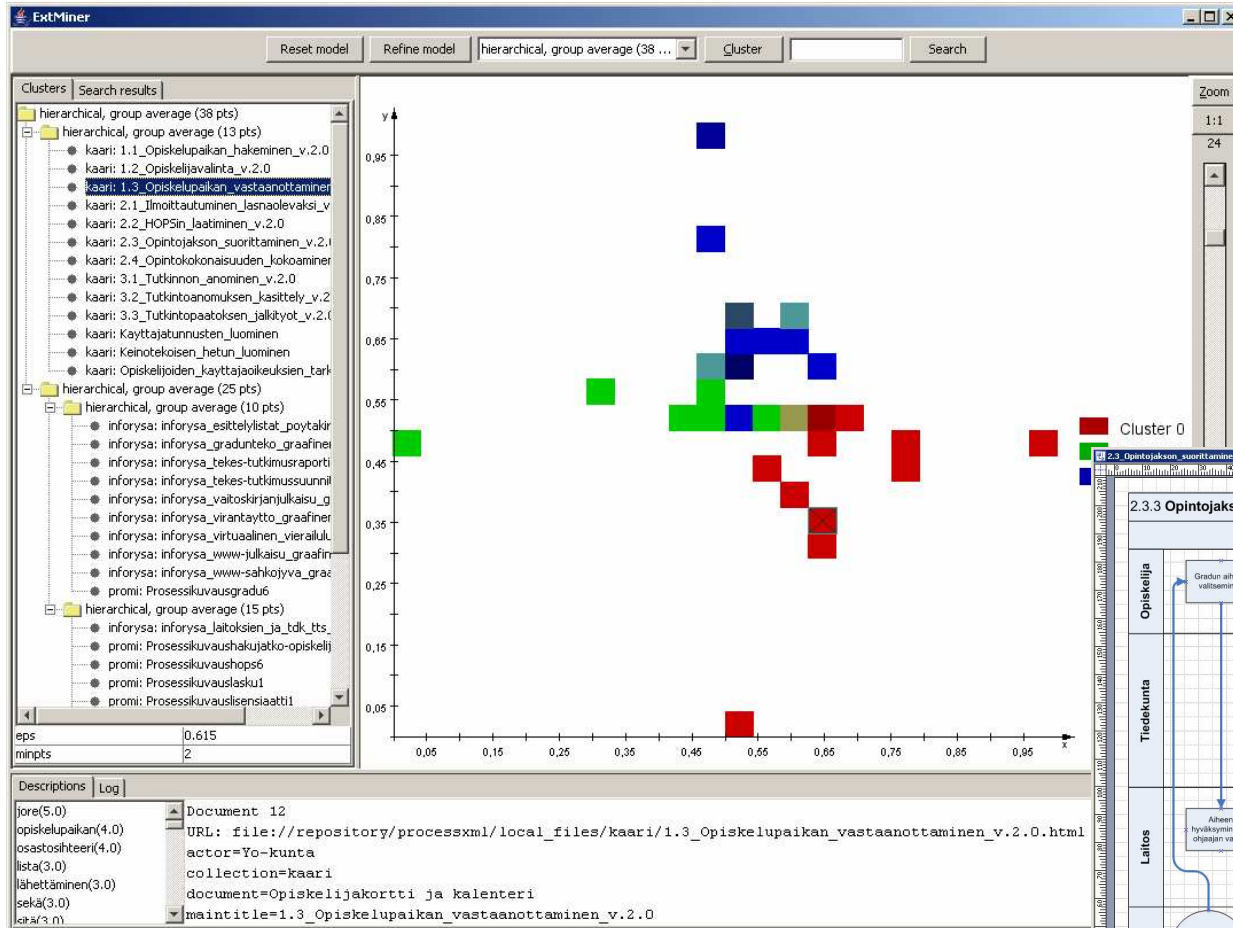
KDD Applied to Content-Based Process Mining

- The **selection** phase involves selecting and converting input model data to a manageable representation that can be consumed by ProcMiner input filters.
- Process model datasets are consolidated to a common representation in the **preprocessing** phase using import filters.
- Process models can be reviewed and modified by the user and **transformed** to ProcML using an export filter. Resulting XML files are input data for ExtMiner.
- In the **data mining** phase, documents representing process models are clustered using ExtMiner. The similarity measure used in searching and clustering is by default the cosine similarity, i.e. the "angle" between the document vectors.
- Clustering results are assessed in the **evaluation** phase. Process clustering produces a new hierarchy or partitioning in addition to decomposition defined by the modeler.

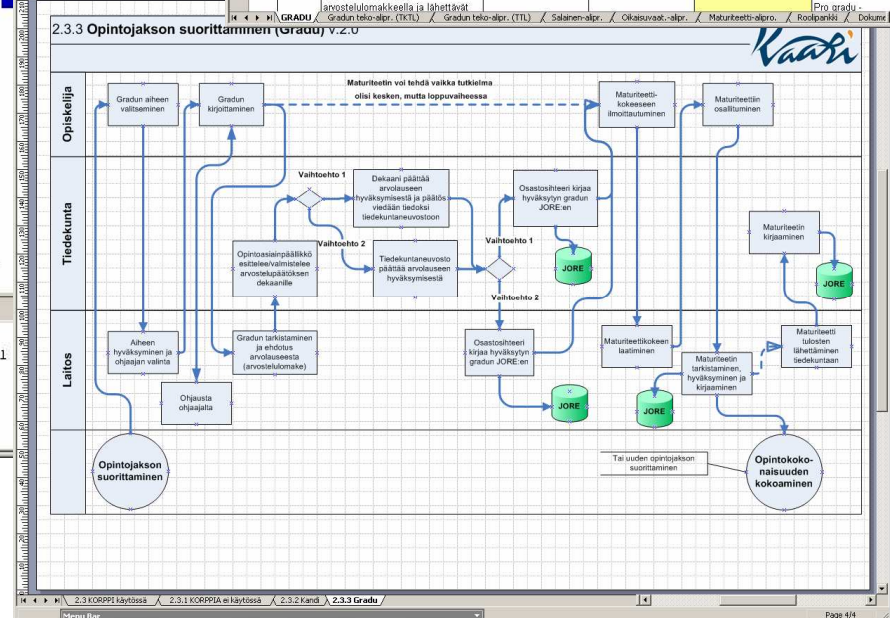
Case 1: Clustering Process Models

- University of Jyväskylä started the implementation of the European quality management initiative at 2005. The Faculty of Information Technology had started modeling their processes on 2001 for developing document management and organizational work.
- To adopt earlier process models to quality system, content-based process clustering was applied to three earlier process modeling projects (38 processes, 167 roles, 178 documents modeled with MS Visio or Excel).
- Process data was consolidated and imported to ProcMiner. The dataset was clustered based on full-text based similarity information using group average hierarchical clustering algorithm.
- It was expected that process clustering would reveal a general topic-based structure, shared by processes modeled by different projects.
- However, the processes were clustered almost entirely according to the original modeling projects.
 - Possible reasons: small number of samples (38) vs features (566 index terms).
 - Subtle differences in terminology and phrasing conventions used in the projects.
 - Hierarchical clustering is affected by the order of documents.

Case 1: Clustering Process Models



Seq #	Prosesin-sinro	VAIHEEN KUVAUS	MITÄ (rooli)	MITÄ (tieto)	MIHIN (rooli)	LOMAKKEET	Linkit/ai-prosessit
1	3.7						Edellä Gradun tekopalrossasi
2	1	Jos TKT:n laitos, niin opiskelija jättää valmiin tutkielman kolmena kappaleena laitoksen kansilaan ja täyttää tarkastajalla saadun sopimuslomakkeen tutkielman laittamisesta tutkielmanpankin www-sivulle	OPISKELIJAA	Tutkielma ja sopimuslomake	LAITOKSEN KANSLIA	Sopimuslomake	Maturiteetti Jos salaiseksi julistettava
3	2	Ohjaaja tekee ehdotuksen kahdeksi tarkastajaksi	OHJAAJAA	Ehdotus tarkastajiksi Tieto tutkielmasta ja esitys tarkastajiksi	AMANUJENSSI		
4	3	Laitoksen johto tutustuu tutkielmaan ja määrää sille kaksi tarkastajaa	AMANUJENSSI		LAITOKSEN JOHTO		Tarkastajien esteellisyydestä
5	4	Jos TKT:n laitos, niin opiskelija jättää tutkielman kolmena kappaleena laitoksen kansilaan ja täyttää sopimuslomakkeen tutkielman laittamisesta tutkielmanpankin www-sivulle	LAITOKSEN JOHTO		AMANUJENSSI		
6	5	Jos TKT:n laitos, niin opiskelija jättää tutkielman kolmena kappaleena laitoksen kansilaan ja täyttää sopimuslomakkeen tutkielman laittamisesta tutkielmanpankin www-sivulle	AMANUJENSSI		LAITOKSEN JOHTO		
7	6	Ohjaaja tekee ehdotuksen kahdeksi tarkastajaksi (jää usein pois, jos kummalakin ohjaajat ovat vähintään FM-tutkinnon suorittaneita)	OHJAAJAA	Ehdotus tarkastajiksi	LAITOKSEN VARAOHTAJAA		Jos salaiseksi julistettava Tarkastajien esteellisyydestä
8	7	Opiskelija lähettää kopiot tutkielmastaan tarkastajille	LAITOKSEN VARAOHTAJAA	Päätös tarkastajista	TARKASTAJAT		
9	8	Tarkastajat antavat lausuntonsa kuukauden kuluessa arastelomakkeella ja lähettävät	LAITOKSEN VARAOHTAJAA	Päätös tarkastajista	OPISKELIJAA		
10	8	Opiskelija lähettää kopiot tutkielmastaan tarkastajille	LAITOKSEN VARAOHTAJAA	Päätös tarkastajista	LAITOKSEN KANSLIA		
11	9	Opiskelija lähettää kopiot tutkielmastaan tarkastajille	OPISKELIJAA	Tutkielma-kopio	TARKASTAJAT		



Case 2: Process Portal

- Parallel to the unsatisfactory process clustering experiment, new processes were modeled manually using ProcML, partially accounting existing process models.
- By Fall 2006, the faculty-specific model database contained 152 process descriptions of different levels (process groups, subprocesses etc), 46 document types, 86 organizational roles and 13 information systems.
- Process portal was used by all project stakeholders including the developer, 3 modelers, steering group, and faculty staff. Public, searchable process repository allowed organization-wide transparent reviews and feedback.
- A "process improvement process" was defined as a part of the other processes, containing guidelines for process modeling, inspection, deviation, and evolution.
- Published process models have proved to be useful as a centralized repository of work instructions and document reference, scattered earlier to different unit-level web pages.
- Process portal and ProcMiner publishing system work as a solid basis for an organization-wide searchable process handbook.

Case 2: Process Portal

```

- <processInstance id="i2" type="sub" state="returned">
  Pro gradu -tutkielman suunnitteleminen
  <actor id="opiskelija" ref="opiskelijaDef" />
  <actor id="gradu_ohjaaja" ref="gradu_ohjaajaDef" />
  <actor id="sv_vastuuprofessori" ref="sv_vastuuprofessoriDef" />
- <abstraction level="0">
- <seq>
  <terminator id="t1" actor="opiskelija" type="start" />
  - <step id="s0" actor="opiskelija" timeconstraint="Pro gradu -seminaarin ensimmäinen vaihe suoritetaan yleensä kolmannen opiskeluvuoden aikana.">
    Opiskelija suorittaa pro gradu -seminaarin ensimmäisen vaiheen tutustuen oman alan pro gradu -tutkielmiin.
    <desc>Suoritus on pakollinen. Vaadittuja esitietoja ovat mm. kandidaattiseminaari ja kandidaatintutkielma</desc>
  </step>
  - <step id="s1" actor="opiskelija">
    Opiskelija valitsee pro gradu -tutkielman aiheen (esimerkiksi laitoksen opettajien avustuksella). Aiheen tulee liittyä johonkin laitoksen suuntautumisvaihtoehtoista.
    <desc>Pro gradu -aiheita on kerätty esimerkiksi Korppi-tietojärjestelmään ja ohjaajien kotisivuille</desc>
  </step>
  <step id="s2" actor="opiskelija">Opiskelija laatii alustavan suunnitelman tutkielman sisällöstä ja aiheen rajauksesta (yhteistyössä laitoksen opettajan) kanssa</step>
  <step id="s3" actor="opiskelija">Opiskelija hyväksyttää tutkielman aiheen ja alustavan suunnitelman oman suuntautumisvaihtoehdonsa vastuuprofessorilla</step>
  <step id="s4" actor="sv_vastuuprofessori">Suuntautumisvaihtoehdon vastuuprofessori hyväksyy aiheen ja kiinnittää pro gradu -tutkielman ohjaajan</step>
  <step id="s5" actor="opiskelija">Opiskelija laatii tarkemman suunnitelman pro gradu -tutkielman sisällöstä, tavoitteista, tutkimusasetelmasta
  
```

Prosessit - Mozilla
[prosessit - rootit - dokumentit - järjestelmät]
Haku

- 2 IT-tiedekunnan asiakasrajapinnan prosessit [html - xml - pdf]
- 3 IT-tiedekunnan toimintaprosessien prosessit [html - xml - pdf]
- 4 Opintoasioiden prosessit [html - xml - pdf]
 - 4.1 Jatko-opiskeluun liittyvät prosessit [html - xml - pdf]
 - 4.2 Perustutkinto-opiskeluun liittyvät prosessit [html - xml - pdf]
 - 4.2.1 Kansainvälistymiseen liittyvät prosessit [html - xml - pdf]
 - 4.2.2 Opetusohjelman hyväksyminen [html - xml - pdf]
 - 4.2.3 Opinto-oppaan laatiminen [html - xml - pdf]
 - 4.2.4 Perustutkinto-opiskelijan elinkaari IT-tiedekunnassa [html - xml - pdf]
 - 4.2.5 Tutkintovaatimusten hyväksyminen [html - xml - pdf]
 - 4.2.6 Opiskelijavalintoihin liittyvät prosessit [html - xml - pdf]
 - 4.2.7 Suuntautumisvaihtoehtoon hakeutuminen [html - xml - pdf]
 - 4.2.8 eHOPSit, siltaopinnot, korvaavuudet [html - xml - pdf]
 - 4.2.9 Pro gradu -tutkielman tekeminen, arvostelu ja julkaisu [html - xml - pdf]
 - 4.2.9.1 Pro gradu -tutkielman tekeminen [html - xml - pdf]
 - 4.2.9.2 Pro gradu -tutkielman tarkastus ja arvostelu [html - xml - pdf]
 - 4.2.9.3 Pro gradu -tutkielman julkaisu [html - xml - pdf]
 - 4.2.10 Perusopiskelijan valmistuminen [html - xml - pdf]
 - 4.3 Sivuaineopiskelijoita koskevat prosessit [html - xml - pdf]

Prosessin pääsekvensi

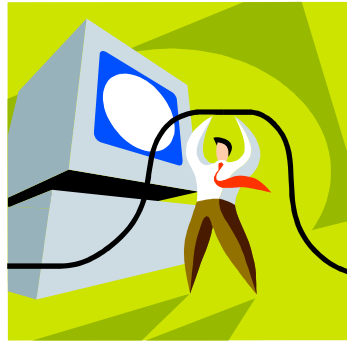
Sekvenssi 1		
Alkupiste	OPISKELIJA	Pro gradu -tutkielman tarkastus ja arvostelu Tulevat linkit [4.2.9.2 Pro gradu -tutkielman tarkastus ja arvostelu]
1.1	OPISKELIJA	Opiskelija täyttää verkkojulkaisusopimuksen, jonka voi tulostaa yliopiston kirjaston verkkosivulla, ja toimittaa sen yliopiston kirjaston julkaisuyksikköön julkaisusihteerille
1.2	OPISKELIJA	Opiskelija tuottaa tutkielmasta PDF-dokumentin esim. yliopiston kirjaston mikrolokaan tai palvelukirjastojen tietokoneilla
1.3	OPISKELIJA	Opiskelija kirjautuu opinnäytteiden julkaisukanavaan ja siirtää PDF-muotoisen tutkielman opinnäytepalvelimelle tai vaihtoehtoisesti toimittaa sen paperilla tai levykkeellä kirjaston julkaisuyksikköön julkaisusihteerille

[Prosessikäsikirja](#)
[Mallitietokanta](#)
[Prosessien väliset linkit](#)

Conclusion and Further Research

- A common object model consolidates process data from diverse sources. ProcML language has been successfully applied for modeling new processes.
- Process retrieval and multichannel publishing simplifies organization-wide applicability and communication of process descriptions both in modeling and implementation stages.
- Structured document clustering may facilitate business process development by providing an independent view to the process subject areas. However, in order to achieve useful clustering results, processes should be modeled using standard, consistent terminology or even based on organizational ontology.
- ProcMiner should be enhanced with additional process consolidation functionality (e.g. detecting multiple connotations inferring to the same actor) and 2-way transforms to facilitate visual process modeling.
- In addition to purely content-based clustering, process data analysis should be based on structural metrics or similarity measures.
- ProcML needs to be cross-analyzed with other process modeling languages.

Thank You!



minurmin@mit.jyu.fi
<http://www.mit.jyu.fi/minurmin/>
<http://extminer.sf.net/>