Cost estimation of Ontologies

Maximilian Hecht & Sascha Pudenz

Applied Ontology Engineering 2010
@ STI Innsbruck
Outline

- Motivation
- ONTOCOM & ONTOCOM II
- ONTOCOM Light
- FOLCOM
- EVEKS Framework
- Conclusion
Motivation

- Rapid developments of the WWW within the last decades
- Large scale dissemination of semantic web and ontologies even into organizational systems
- Economical view on development gets more and more important

The *effort (costs) of ontology development projects* needs to be estimated to enable a evaluation from a cost/benefit perspective!
ONTOCOM I

- **Ontology Cost Model** was introduced in 2005
- Estimates the overall effort of ontology development in person months (PM) parametrically

**Estimation:**
- Size of the ontology estimated in thousand ontology primitives
  - e.g. concepts, relations, axioms and instances
- $\alpha$ represents non-linear behavior of the model
- $A$ represents size independent costs
- $CD$ are cost drivers

$$PM = A \times \text{Size}^\alpha \times \prod CD_i$$
ONTOCOM I – Cost Drivers

- Predefined
- Have a rating level and related scales (from Very Low to Very High)
- Grouped in:
  - Product cost drivers
  - Personnel cost drivers
  - Project cost drivers
- Example:
  - Documentation Complexity from Very Low to Very High

\[ PM = A \times \text{Size}^\alpha \times \prod CD_i \]
ONTOCOM I Example

- Tool by STI Innsbruck

![Ontology Size Diagram](image-url)
ONTOCOM II

Process-based framework
1. Acquire and understand domain knowledge
2. Analyze functional requirements of development project
3. Estimate the effective size of the ontology
4. Estimate ontology development efforts
5. Estimate ontology reuse and maintenance efforts
6. Analyze available personnel supports
7. Estimate project and schedule efforts

Other small improvements
- More precise estimation of Size
- New grouping of cost drivers
ONTOCOM Lite

- **ONTOCOM Lite** for the domain of taxonomies (lightweight ontologies) as a variant of ONTOCOM
  - Taxonomy development cycle as first basis to identify cost drivers:

  ![Taxonomy Development Cycle Diagram]

  - **Cost driver groups** similar to ONTOCOM, however,
    - some cost drivers were removed due to non-applicability
    - two additional cost driver groups to cover usage drivers
      - Reuse cost drivers (e.g. Taxonomy Understanding or Taxonomy Translation)
      - Usage cost drivers (e.g. Complexity of the user feedback or reporting)
  - Calculation analogous to ONTOCOM [4,6]
FOLCOM

FOLCOM: Cost Estimation Model for Folksonomies [5]
- basic principles derived from the story points method used in agile software estimation
- Assumption: users construct a folksonomy by a sequence of three main operations: add, remove and modify annotations
- Costs for optional operations and infrastructure costs are not considered → estimate separately.
- Effort is estimated in ideal time = amount of time that something takes without peripheral activities

4-step approach:
1. Requirements analysis
2. Size estimation
3. Velocity determination
4. Effort estimation
FOLCOM (contd.)

Requirements Analysis

Check if requirements (certain assumptions) hold
Possibility to measure tagging times from users?
Single or multi-tagging? How many objects and what objects?...

Size Estimation

Calculate the number of Story Points $sp$
Each object in the object $Collection := \{O_1, O_2, \ldots, O_n\}$ represents a tagging task. The size of each task is estimated by assigning a numerical size/complexity value

$$sp_{col} := \sum_{i=1}^{n} \text{complexity}(O_i)$$

Velocity Determination

Calculate velocity by tagging some objects of the collection; e.g. 2min/sp

$$velocity := \frac{\sum_{i=1}^{n} \text{taggingTime}(O_i)}{\sum_{i=1}^{n} \text{timesTagged}(O_i)}$$

$$multiTagFactor := \frac{\sum_{i=1}^{n} \text{time} \times sp_{it}}{\sum_{i=1}^{n} \text{complexity}(O_i)}$$

Effort Estimation

Calculate Effort for remaining story points

$$\text{effortEstimation}_{rem} := \text{multiTagFactor} \times sp_{rem} \times velocity$$

Note: In single tagging projects $multiTagFactor = 1$
EVEKS Framework

- Enterprise Value Estimation for Knowledge-based Systems
- Scope: effective prediction of development effort and cost/benefits of knowledge-based applications under one umbrella
- Steps inside the EVEKS framework:

![Diagram of EVEKS Framework](image-url)

- ONTOCOM
- ONTOCOM II
- FOLCOM
- ONTOCOM LITE...
Conclusion

- Complete and calibrated methods:
  - ONTOCOM (I+II) for cost estimation of ontology projects
  - FOLCOM for cost estimation of folksonomy projects
- Not yet calibrated but with potential:
  - ONTOCM Lite for taxonomy projects
- EVEKS: integrating framework for knowledge-based applications development projects providing a cost/benefit perspective

A powerful tool set that helps to plan and estimate ontology based projects

Improves managerial decision making and justification of investments for knowledge-based applications within organizations
Thank you for your attention!

References:


