Important concepts in CMMI
and
what is difficult to understand

Dr. Katalin Balla
Zádor Dániel Kelemen

Budapest University of Technology and Economics
SQI Hungarian Software Quality Consulting Institute Ltd.

JUNE 7-9, 2011
Dublin, Ireland
Acknowledgement

• Preparing and delivering this presentation was sponsored by
  • BUTE Centre of Information Technology
    • BME Információtechnológiai Innovációs és Tudásközpont,
      http://www.ik.bme.hu/index_eng.php

  within project SZOMINo8
  TECH_o8_A2-SZOMINo8 - Software quality assurance service-package for open document format applications

  AND

  within project „Q-Works” - GOP-1.1.1-08/1-2008-0023 – Automated workflow system for quality assurance, based on CMMI
Contents

- Introduction
  - Ways of understanding CMMI
- Text mining in CMMI
  - What we did
  - Results of text mining
- Difficulties in applying CMMI concepts
  - What people frequently misunderstand
    ✦ Based on work with companies implementing CMMI
Introduction -1

- CMMI: huge amount of text
  - Understandable by
    ✦ Reading it / interpreting it, using experience and knowledge related to SPI, concepts of quality, other models and standards
    ✦ Text mining
    ✦ ...

Important concepts in CMMI and what is difficult to understand
Introduction -2

Understanding CMMI

- By reading it / interpreting it, using experience and knowledge related to SPI, concepts of quality, other models and standards
- PhD research of Zádor Dániel Kelemen. Poster presentation on SEPG 2009, Prague.

**Elements:** Process area, Specific Goal, Specific Practice, Typical Work Product, Subpractice, Generic Goal, Generic Practice, Generic Practice Elaboration
Contents

- Introduction
  - Ways of understanding CMMI
- Text mining in CMMI
  - What we did
  - Results of text mining
- Difficulties in applying CMMI concepts
  - What people frequently misunderstand
    ✦ Based on work with companies implementing CMMI
Understanding the most important concepts of CMMI

• “In God we trust; all others must bring data.” – W. Edwards Deming.
  ▪ We got data from SEI: ~3*500 pages
  ▪ Somehow we wanted to process it
  ▪ Tools used:
    ✦ open source text mining tool: RapidMiner
    ✦ Excel,
    ✦ MySQL,
    ✦ ...and a friend, Imteaz for manual verification
The most important concept
The most important concept

- All occurrences printed would make a 1627 meters long line
- 20 minutes walking distance (Google maps)
- More than
  - walking from Clark Adams square - center of Hungary
  - passing the Danube Europe’s second largest river - on Chain bridge
  - walking till the Holy Crown of Hungary - which can be found inside the Hungarian Parliament
Processing CMMI documents

3 simple steps:
1. Processing documents from pdf/word files
2. Converting wordlist to data
3. Writing results to an output file (e.g. excel)
Processing documents

Important concepts in CMMI and what is difficult to understand

8 June 2011
<table>
<thead>
<tr>
<th>#</th>
<th>word</th>
<th>in documents</th>
<th>total</th>
<th>#</th>
<th>word</th>
<th>in documents</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>process</td>
<td>3</td>
<td>8946</td>
<td></td>
<td>process</td>
<td>3</td>
<td>10853</td>
</tr>
<tr>
<td>2</td>
<td>work</td>
<td>3</td>
<td>3706</td>
<td></td>
<td>product</td>
<td>3</td>
<td>4370</td>
</tr>
<tr>
<td>3</td>
<td>project</td>
<td>3</td>
<td>3170</td>
<td></td>
<td>servic</td>
<td>3</td>
<td>4219</td>
</tr>
<tr>
<td>4</td>
<td>service</td>
<td>3</td>
<td>2934</td>
<td></td>
<td>work</td>
<td>3</td>
<td>3751</td>
</tr>
<tr>
<td>5</td>
<td>cmmi</td>
<td>3</td>
<td>2682</td>
<td></td>
<td>project</td>
<td>3</td>
<td>3556</td>
</tr>
<tr>
<td>6</td>
<td>management</td>
<td>3</td>
<td>2532</td>
<td></td>
<td>perform</td>
<td>3</td>
<td>3501</td>
</tr>
<tr>
<td>7</td>
<td>performance</td>
<td>3</td>
<td>2437</td>
<td></td>
<td>manag</td>
<td>3</td>
<td>3459</td>
</tr>
<tr>
<td>8</td>
<td>requirements</td>
<td>3</td>
<td>2406</td>
<td></td>
<td>requir</td>
<td>3</td>
<td>3022</td>
</tr>
<tr>
<td>9</td>
<td>product</td>
<td>3</td>
<td>2338</td>
<td></td>
<td>plan</td>
<td>3</td>
<td>2988</td>
</tr>
<tr>
<td>10</td>
<td>organization</td>
<td>3</td>
<td>2194</td>
<td></td>
<td>area</td>
<td>3</td>
<td>2930</td>
</tr>
<tr>
<td>11</td>
<td>area</td>
<td>3</td>
<td>2044</td>
<td></td>
<td>cmmi</td>
<td>3</td>
<td>2682</td>
</tr>
<tr>
<td>12</td>
<td>products</td>
<td>3</td>
<td>1903</td>
<td></td>
<td>organ</td>
<td>3</td>
<td>2546</td>
</tr>
<tr>
<td>13</td>
<td>processes</td>
<td>3</td>
<td>1879</td>
<td></td>
<td>includ</td>
<td>3</td>
<td>2319</td>
</tr>
<tr>
<td>14</td>
<td>organizational</td>
<td>3</td>
<td>1641</td>
<td></td>
<td>measur</td>
<td>3</td>
<td>2124</td>
</tr>
<tr>
<td>15</td>
<td>information</td>
<td>3</td>
<td>1589</td>
<td></td>
<td>risk</td>
<td>3</td>
<td>2089</td>
</tr>
<tr>
<td>16</td>
<td>version</td>
<td>3</td>
<td>1577</td>
<td></td>
<td>develop</td>
<td>3</td>
<td>2017</td>
</tr>
<tr>
<td>17</td>
<td>objectives</td>
<td>3</td>
<td>1545</td>
<td></td>
<td>establish</td>
<td>3</td>
<td>1969</td>
</tr>
<tr>
<td>18</td>
<td>include</td>
<td>3</td>
<td>1538</td>
<td></td>
<td>improv</td>
<td>3</td>
<td>1924</td>
</tr>
<tr>
<td>19</td>
<td>analysis</td>
<td>3</td>
<td>1366</td>
<td></td>
<td>exampl</td>
<td>3</td>
<td>1863</td>
</tr>
<tr>
<td>20</td>
<td>supplier</td>
<td>3</td>
<td>1359</td>
<td></td>
<td>object</td>
<td>3</td>
<td>1798</td>
</tr>
<tr>
<td>21</td>
<td>data</td>
<td>3</td>
<td>1298</td>
<td></td>
<td>inform</td>
<td>3</td>
<td>1769</td>
</tr>
<tr>
<td>22</td>
<td>services</td>
<td>3</td>
<td>1285</td>
<td></td>
<td>supplier</td>
<td>3</td>
<td>1714</td>
</tr>
<tr>
<td>23</td>
<td>training</td>
<td>3</td>
<td>1274</td>
<td></td>
<td>organiz</td>
<td>3</td>
<td>1650</td>
</tr>
<tr>
<td>24</td>
<td>development</td>
<td>3</td>
<td>1262</td>
<td></td>
<td>level</td>
<td>3</td>
<td>1638</td>
</tr>
<tr>
<td>25</td>
<td>quality</td>
<td>3</td>
<td>1261</td>
<td></td>
<td>identifi</td>
<td>3</td>
<td>1636</td>
</tr>
<tr>
<td>26</td>
<td>risk</td>
<td>3</td>
<td>1225</td>
<td></td>
<td>use</td>
<td>3</td>
<td>1603</td>
</tr>
<tr>
<td>27</td>
<td>plan</td>
<td>3</td>
<td>1215</td>
<td></td>
<td>version</td>
<td>3</td>
<td>1594</td>
</tr>
<tr>
<td>28</td>
<td>activities</td>
<td>3</td>
<td>1203</td>
<td></td>
<td>select</td>
<td>3</td>
<td>1567</td>
</tr>
<tr>
<td>29</td>
<td>level</td>
<td>3</td>
<td>1113</td>
<td></td>
<td>practic</td>
<td>3</td>
<td>1549</td>
</tr>
</tbody>
</table>
The most important concepts in QMIM

- Process
- Product
- Work/Project
- Service

Important concepts in CMMI and what is difficult to understand
Finding cross-references in CMMI

Conducting the search

• We were looking for:
  ▪ “Refer to the” ... “process area ...”
  OR
  ▪ “Refer to the” ... “specific practice in” ... “process area...”

• Part 2 of all constellations were examined

• Examples were filtered

• Irrelevant references were filtered, e.g.:
  ▪ the organization’s set of standard processes can refer to the standard processes established at the organization level
Cross-references in CMMI

Results:
- Quite complex: 992 (1016) cross-references in total!
  - 311 in CMMI-DEV
  - 388 in CMMI-SVC
  - 293 in CMMI-ACQ
- Referring from different levels:
  - In PAs: Introductory notes, related process areas, specific practice level
  - In GPs
- Referring to different levels:
  - E.g. to PA, SG, SP
Understanding Cross-references in CMMI-DEV (referring from)

Training: GPs and PPQA are probably easier to understand

IPM implementation can be quite difficult (especially in multimodel environment),
since it relies on many other elements

Important concepts in CMMI and what is difficult to understand
Understanding Cross-references in CMMI-DEV (referring to)

Implementing MA, PP, PMC will have effect on many other elements

PPQA, CAR, OT can probably be implemented more independently.

What about PI?

Important concepts in CMMI and what is difficult to understand
This is a network! having ~Pareto distribution! (only 16.6% of elements have more than 6 interconnections with one another element)

10 from PMC to PP, 1 from PP to PMC -> wise to implement them together

PPQA is the most isolated
TS-RD (8,4) -> 12; IPM-OPD (8,1); OPM-OPD (6,2); QPM-OPD (4,3);
Views of the network
Views of the network
Views of the network
Visualizing characteristics

NetworkAnalyzer - Visualize Parameters

Apply visual styles to **Connections.2**

Map node size to:
- **EdgeCount**
  - Low values to small sizes
  - Low values to large sizes

Map node color to:
- **NeighborhoodConnectivity**
  - Low values to bright colors
  - Low values to dark colors

Map edge size to:
- **EdgeBetweenness**
  - Low values to small sizes
  - Low values to large sizes

Map edge color to:
- **EdgeBetweenness**
  - Low values to bright colors
  - Low values to dark colors

[Image of the NetworkAnalyzer interface]
Visualizing characteristics
Contents

- Introduction
  - Ways of understanding CMMI
- Text mining in CMMI
  - What we did
  - Results of text mining
- Difficulties in applying CMMI concepts
  - What people frequently misunderstand
    ✦ Based on work with companies implementing CMMI
Difficult to understand for CMMI users…

- (Top 5 - based on own experience)…
  - Planning and monitoring activities different from engineering ones
  - CM for things different from code
  - Data management – often confused with CM
  - IPM as a whole (often argued that if PP, PMC is done for all projects, IPM results – so, e.g. no process description is needed)
  - GP-s, mainly planning of PP and PMC, measurement on measurement, CM on CM, relevant stakeholders for PA-s …
Some statistics

Weaknesses and notes referring to CMMI ML2 PA-s
(based on 9 SCAMPI A-s, all weaknesses+notes = 100%, 2007-2010)

- **PP** 30%
- **PPQA** 4%
- **PM** 10%
- **CM** 21%
- **MA** 20%
- **SAM** 6%
- **REQM** 9%

Most weaknesses are in connection with PP and PMC.

Most of weaknesses refer to "maintain bidirectional traceability".

Measurement has little feedback. More PA-s measured "together".

Configuration management, and mainly configuration audits is sometimes formal for work products different from code.
Some statistics

Weaknesses and notes referring to CMMI ML3 PA-s
(Excl. SAM, 8 SCAMPI A-s, 2008-2010. All weaknesses + notes = 100%)

3 PA-s out of 17 PA-s 3 have 41% of all weaknesses and notes (PP+PMC+IPM)
Some statistics

Strengths for ML3 PA-s
(9 SCAMPI A-s, total nr. of strengths=100%)

- RD, REQM: 14%
- PMC: 14%
- IPM: 14%
- PP: 22%
- Other: 36%
- "Other" more detailed:
  - TS: 14%
  - CM, PI, OT: 21%
  - Other PA-s

Important concepts in CMMI and what is difficult to understand
Some statistics

Average of ML2 PA SP-conformity for pre-appraisals
(15 pre-appraisals, PA SP conformity = 100%)

- PP: 87%
- PMC: 83%
- CM: 94%
- MA: 76%
- REQM: 78%
- SAM: 95%
- PPQA: 87%

Important concepts in CMMI and what is difficult to understand
What we talked about…

• What is important in CMMI – based on text analysis
  ▪ Confirms our previous research

• How process areas are interconnected – based on text analysis
  ▪ Results could be useful to help companies use continuous representation

• What are the most difficult and misunderstood concepts when implementing CMMI – based on historical data
Thank you for your attention!

balla.katalin@sqi.hu
kelemen.daniel@sqi.hu
www.sqi.hu
www.iit.bme.hu