What are the used UML diagrams?  
A Preliminary Survey

Gianna Reggio, Maurizio Leotta, Filippo Ricca, Diego Clerissi

DIBRIS – Università di Genova, Italy
Our aim

• Asses by means of a survey which parts of the UML (diagrams and constructs) are the most/less used in practice
  – By examining “objective” sources
    • books
    • university IT courses
    • tutorials
    • tools
  – By a personal survey aimed to UML users
Motivation (1)

• UML is really BIG (and complicate)
  – 14 different diagrams
    • large set of constructs for each diagram
  – huge specification book (732 pages)
  – large and complex metamodel
  – not easy to understand its static and dynamic semantics

• As a consequence
  – it is difficult to teach
  – to define UML model transformations is time consuming and complex
  – sometime simple ad hoc lean DSLs are preferred
Motivation (2)

• On the other hand
  – users naturally tend to consider and use only a part of the UML
    – “For 80% of all software only 20% of UML is needed” I. Jacobson on his blog
  – people usually learn the UML on books/courses/tutorials using a model editor tool, that in general do not cover the whole UML
    – thus, they will never become aware of the existence of many specific constructs
Motivation (3)

• Why bother if a large part of the UML is scarcely used, if not almost unknown
  – can cause a waste of effort and of resources by who want/must use UML
    • trivially printing the reference requires 700 sheets
    • understanding the metamodel / preparing for the certification / deciding what to teach to the students / reading a UML book require a large number of hours
    • maintaining the official specification requires a large amount of effort
  – Also the OMG has recently recognised the need to simplify the UML with the initiative “UML Simplification”, but it concerns only the way UML is defined
The survey

• To assess which parts of the UML (diagrams/constructs) are the most/less used in practice
  – Investigating objective sources
    • books about the UML
    • IT University courses covering also the UML
    • tutorials presenting the UML to practitioners
    • UML model editors
  – By a personal opinion survey asking to UML users of different kinds which parts of the UML they know

• Current situation
  – Done for diagrams, and activity and use case diagram constructs for objective sources
    • Today the results concerning the UML diagrams
  – The personal survey* starts now

* restricted to the same items (diagrams, and activity and use case diagram constructs)
Study definition

• Main research question

  Which of the 14 types of UML diagrams are the most/less used in practice?

• Target population

  – “Objective” sources about UML
    • books, tools, courses and tutorials
  – Source selection
    • Conducted a systematic search performed using Internet resources, Web search engines and electronic databases
    • used non-probabilistic (convenience sampling) methods
    • adopted some inclusion/exclusion criteria

• Data extraction

• Analysis performed the using descriptive statistics
Inclusion/Exclusion Criteria

• Common to all sources
  – **UML versions ≥ 2.0**

• Books
  – Last edition (whenever possible)
  – No gray literature (books without ISBN)

• Tools
  – UML modelling tools (commercial and non-commercial)
  – Excluded
    • general graphics editor
    • tools providing only a specific type of diagram
    • unstable, not complete or preliminary tools

• Courses
  – only university courses concerning IT
  – also not in English (e.g., Italian, Spanish)

• Tutorials
  – common meaning/perception
    more interactive and specific than a book or a lecture, seeking to teach by examples
Source collection (1)

- Books
  - Amazon website
    - different search criteria (most results using “UML 2”)
  - filtered out books not satisfying the inclusion criteria
  - tried to recover them
  - At the end 30 books (18 are in the top 24 ranked by relevance by Amazon), e.g.,
      “The Unified Modeling Language User Guide”
      by J. Rumbaugh, I. Jacobson, G. Booch
    - “UML Distilled” by M. Fowler.
Source collection (2)

• Tools
  – Used
    • Wikipedia “List of Unified Modeling Language tools” (49 UML tools)
    • UML-tools web-site www.uml-tools.com
    • Google searches using different string combinations
  – Checked each tool w.r.t. the exclusion criteria
  – Down-loaded and installed the most recent versions
  – At the end 20 different tools, e.g
    • Visual Paradigm, MagicDraw, IBM Rational SW Architect
Source collection (3)

• Courses
  – started using Google
  – found several university courses satisfying the inclusion criteria stated above, but no way to find the lecture slides
  – resort also to convenience sampling, asking to our colleagues
  – At the end, 22 different University courses, from different countries (Canada, UK, USA, Hungary, Germany, Italy, France, Spain, Argentina, Australia)

• Tutorials
  – Started with three lists
  – Google search
  – At the end 18 tutorials
Survey results

Preliminary assumptions

- A diagram is
  - **widely used** if it is present in the $\geq 60\%$ of the sources
  - **scarcely used** if it is present in $\leq 40\%$ of the sources,
  - non-defined in the other cases (grey zone)

- Books classified in
  - Guide: concerned only on the UML (15 books)
  - Spec: use of the UML for specific tasks (15 books)

- “Used” meaning
  - For tools
    - a diagram is used if we can produce a model containing it
  - For books, courses, tutorials
    - A diagram is used if it is mentioned
      No way to find a better definition
Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Scarcely used

- All not present in UML 1.x
- Profile diagram
  - only from UML 2.2,
    - (essentially a variant of package diagram)
- Timing
  - restrict scope,
    - UML offers other ways to model time related aspects
### Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

**Widely used**

- Class diagram 100%
- Activity diagram 98%, perhaps because used for business processes and SOA
- All, except package, present in UML 1.x (but communication diagram was called collaboration diagram)
Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

• Composite structure:
  - Appearing in UML 2.0
  - Surprising because structured classes were filling a hole
  - Almost two different things: collaborations and structured classes
### Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

#### Books

- **Guide**
  - All diagrams, except profile, widely used
  - Perhaps due to the fact that we consider used a diagram if it is just mentioned
Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>95%</td>
<td>95%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

- Scarcely used: only Timing and Profile diagram
- Grey area: Interaction overview

Surprising, but

- if you can draw a class diagram it is quite easy to handle composite structure, component, deployment, object, package, communication and profile diagrams
- whereas timing and overview requires new graphical functionalities
Survey results

- Striking differences
  - Widely used almost 100%: **class, activity, sequence, use case, state machine**
  - Grey area: Deployment, object, package, communication, component
  - Scarcely used: composite structure 14%, Interaction overview, timing, profile ≤5%
- But a lecturer has to decide which are the most relevant diagrams to comply with the course constraints
  - this decision is quite homogeneous
### Survey results

<table>
<thead>
<tr>
<th>UML Diagram</th>
<th>Book Guide</th>
<th>Book Spec</th>
<th>Tool</th>
<th>Course</th>
<th>Tutorial</th>
<th>All Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Composite Structure</td>
<td>87%</td>
<td>56%</td>
<td>80%</td>
<td>14%</td>
<td>33%</td>
<td>52%</td>
</tr>
<tr>
<td>Component</td>
<td>93%</td>
<td>75%</td>
<td>85%</td>
<td>59%</td>
<td>89%</td>
<td>79%</td>
</tr>
<tr>
<td>Deployment</td>
<td>93%</td>
<td>81%</td>
<td>90%</td>
<td>55%</td>
<td>89%</td>
<td>80%</td>
</tr>
<tr>
<td>Object</td>
<td>93%</td>
<td>81%</td>
<td>70%</td>
<td>55%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Package</td>
<td>100%</td>
<td>80%</td>
<td>65%</td>
<td>52%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>Activity</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Sequence</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>100%</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Communication</td>
<td>100%</td>
<td>81%</td>
<td>90%</td>
<td>59%</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>Interaction Overview</td>
<td>80%</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
<td>28%</td>
<td>38%</td>
</tr>
<tr>
<td>Timing</td>
<td>87%</td>
<td>50%</td>
<td>40%</td>
<td>5%</td>
<td>33%</td>
<td>40%</td>
</tr>
<tr>
<td>Use Case</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>State Machine</td>
<td>100%</td>
<td>94%</td>
<td>100%</td>
<td>95%</td>
<td>89%</td>
<td>96%</td>
</tr>
<tr>
<td>Profile</td>
<td>7%</td>
<td>13%</td>
<td>30%</td>
<td>0%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

- **Tutorials**
  - Same trend of courses, but less striking differences
  - Also a tutorialist has to decide what present, but it seems s(he) has less stringent constraints
Threats to Validity

• Only sources aimed to UML, e.g., no
  – drawing tools able to produce pictures of UML diagrams
  – books presenting a survey of visual notations

• Considered all possible sources?
  – Yes: books, tools (100% of existing ones), tutorials, courses
  – No: Research papers (too biased versus peculiarities, problems)
  – No: generic web resources (too difficult to evaluate them)
  – **UML users**: will be yes, we are launching now the online survey


• Computed widely/scarcely used all sources, disregarding their quite different kinds
  No way to assign to them different weights in an unbiased way
Conclusions

• Investigated, by means of a survey, how much the UML diagrams are used
  – considering “objective” sources: books, tools, courses, and tutorials
  – checking if a diagram it is used in an objective way, so these results are not biased by any personal opinion

• An “essential” UML seems to emerge also if its boundaries are not exactly defined
  – widely used without any doubts: class, activity, sequence, use case and state machine
  – really scarcely used: interaction overview, timing and profile diagrams

• Combined with the results of the just starting personal survey should lead to finally determine an “essential” UML
Take part in the online survey


Thanks!