Journal of Graph Algorithms and Applications
http://www.cs.brown.edu/publications/jgaa/
vol. 6, no. 3, pp. 255–279 (2002)

Graph Layout Aesthetics in UML Diagrams: User Preferences

Helen C. Purchase *Jo-Anne Allder* *David Carrington*

School of Information Technology and Electrical Engineering
The University of Queensland
Brisbane 4072, Australia
{hcp,joanne,davec}@itee.uq.edu.au

Abstract

The merit of automatic graph layout algorithms is typically judged by their computational efficiency and the extent to which they conform to aesthetic criteria (for example, minimising the number of crossings, maximising orthogonality). Experiments investigating the worth of such algorithms from the point of view of human usability can take different forms, depending on whether the graph has meaning in the real world, the nature of the usability measurement, and the effect being investigated (algorithms or aesthetics). Previous studies have investigated performance on abstract graphs with respect to both aesthetics and algorithms, finding support for reducing the number of crossings and bends, and increasing the display of symmetry.

This paper reports on preference experiments assessing the effect of individual aesthetics in the application domain of UML diagrams. Subjects' preferences for one diagram over another were collected as quantitative data. Their stated reasons for their choice were collected as qualitative data. Analysis of this data enabled us to produce a priority listing of aesthetics for this domain. These UML preference results reveal a difference in aesthetic priority from those of previous domain-independent experiments.

Communicated by Michael Kaufmann: submitted April 2001;
revised December 2001 and June 2002.

1 Introduction

The success of automatic graph layout algorithms which display relational data in a graphical form is typically measured by their computational efficiency and the extent to which they conform to aesthetic criteria (for example, minimising the number of crossings, maximising orthogonality). In addition, designers of these algorithms often claim that by conforming to these aesthetic criteria, the resultant graph drawing helps the human reader to understand the information embodied in the graph. However, little research has been performed on the usability aspects of such algorithms: do they produce graph drawings that make the embodied information easy to use and understand? Is the computational effort expended on conforming to conventional aesthetic criteria justifiable with respect to better usability? As automatic graph layout algorithms are increasingly being used in information visualisation systems (for example, for the visualisation of social networks or data-flow diagrams), it is important that the effectiveness of these algorithms from a user point of view is investigated, to ensure that the algorithms used are appropriate for the domain being modelled.

In defining a framework for experimentation in this area, we identify three dimensions of usability studies investigating the merit of graph drawing algorithms from a human perspective: the nature of the graph (syntactic or semantic), the type of usability measurement (preference or performance), and the effect being investigated (algorithms or aesthetics).

- *Syntactic* graph drawing experiments use a graph structure that has no meaning in the real world: it is merely an abstract collection of nodes with relationships between them. *Semantic* graph drawing experiments use a graph within a particular application domain: in this case, the graph has meaning in the real world (for example, a transport network, or a data-flow diagram), and the experimental tasks are performed in relation to the information represented.

- *Preference* experiments ask the subjects to state their preference of one drawing over another. Subjects may or may not be able to explain the reasons for their preference: sometimes it may be based on an inexplicable personal view. A more quantifiable measurement of usability can be determined in *performance* experiments, where subjects are required to perform a particular task (or tasks) using a given graph: the data collected in performance experiments is the extent of the subjects' success in performing the task.

- Two possible effects on usability may be investigated in graph drawing experiments: the effect of individual *aesthetics* (e.g. reducing crossings, maximising orthogonality) and the effect of the use of different *algorithms* (producing drawings conforming to different aesthetics to varying degrees). In the case of investigating aesthetics, the experimental drawings need to be produced by hand, with appropriate manipulation of the variables; in the case of investigating algorithms, the experimental drawings would be produced by existing layout algorithms,

Two previous studies have investigated syntactic performance. The first experiments considered the effect of individual aesthetics, and found support for reducing the number of crossings and bends, and increasing the display of symmetry. However, no support was found for maximising the minimum angle or increasing orthogonality (Purchase, 1997). The second experiment considered the effect of eight different algorithms, and revealed that it is difficult to say that one algorithm is 'better' than another in the context of syntactic understanding of the abstract graph structure (Purchase, 1998).

This paper reports on two semantic, preference experiments that investigated the effect of individual aesthetics. The application domains are the presentation of two types of UML diagrams: class diagrams and collaboration diagrams. Performance experiments are left for a later study.

All these experiments are part of a larger project, the aim of which is to perform a thorough empirical investigation of the aesthetics underlying graph layout algorithms, and the algorithms themselves, in an attempt to influence the future design of graph layout algorithms through empirical human (rather then computational) experimentation.

2 Experimental scope and definition

2.1 The application domain: UML diagrams

Many different methods and models have been proposed to capture a complete specification of requirements and a comprehensive design representation in a formal software engineering process.

The Unified Modeling Language (UML) (Booch, Rumbaugh and Jacobson, 1998) was chosen as the semantic domain for these preference, aesthetics experiments, as it is rapidly becoming the de facto standard for modelling object oriented systems. UML provides a mainly graphical notation to represent the artifacts of software systems. The notation is relatively new but it is rapidly being adopted as the accepted notation for object-oriented analysis and design. UML incorporates notations to describe systems at various levels of abstraction. UML diagrams can be used to model requirements, designs, implementations and tests. Since these diagrams are a means of communication between customers, developers and others involved in the software engineering process, it is critical that the notation is standardised, and that the diagrams present information clearly. Appropriate layout of these diagrams can assist in achieving this goal.

UML uses several different types of graph drawings that aim to describe a system to meet the users' needs at reasonable cost. Two UML diagram types, class and collaboration, were selected for the experiments reported here.

Class diagrams describe the types of objects in the system, and the static relationships between them. These relationships are either subtypes (representing inheritance) or associations (representing other types of relationship).

Collaboration diagrams indicate how the objects in a system collaborate. They show the interaction of objects and the sequence of events by numbering the events in the order in which they occur (along the arcs), referring to objects as nodes in the graph.

While other studies (for example Irani and Ware, 2000) address user performance with UML, they do not do so from the perspective of automatic graph layout algorithms and aesthetics.

2.2 Aims

The aim of the experiment is to identify an ordered list of the aesthetic features preferred by subjects when embodied in UML class and collaboration diagrams. Such a list can indicate to interface designers of CASE tools the most suitable way to lay out their diagrams for the best response from users. It will also form the basis for more extensive experiments concentrating on performance of users in software engineering tasks.

The study aimed to identify the 'subjectively pleasing' aesthetics in graph drawings. Evaluation of the graph drawings was done solely according to human, individual preference. Performance with respect to a task or correctness of interpretation has been left for a further study.

2.3 Aesthetics investigated

Using graphs from a semantic domain instead of an abstract graph structure introduces additional secondary notations that are particular to the formal semantic notation. Secondary notations are layout or graphical cues that tend not to be part of the formal notation (e.g. adjacency, clustering, white space) (Petre, 1995). Thus, while this experiment included some graph drawing aesthetics as advocated by designers of generic layout algorithms, it also included investigation of other layout features specifically related to the standard UML notation.

For each experiment (class and collaboration), a suitable subset of aesthetic features was identified. These were selected based on emphases in the literature (e.g. Colman and Stott Parker (1996), and Petre (1995)), and as ones that could feasibly be applied to UML diagrams. Many of them were also considered in the prior experiments (Purchase 1997, 1998) and most could be used for both types of diagram.

Six aesthetics were evaluated for both class and collaboration diagrams:

- *minimize bends* (the total number of bends in polyline edges should be minimized (Tamassia, 1987))
- *minimize edge crossings* (the number of edge crossings in the drawing should be minimized (Reingold and Tilford, 1981))

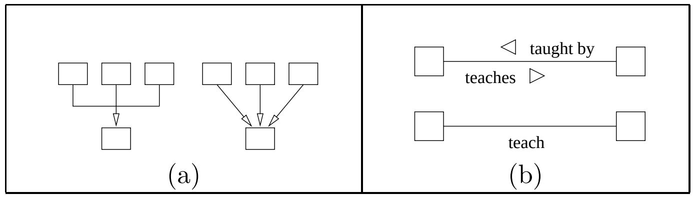

Figure 1: The UML class diagram secondary notation features investigated, showing both alternatives: (a) inheritance, (b) directional indicators.

- *orthogonality* (fix nodes and edges to an orthogonal grid (Tamassia, 1987; Papakostas and Tollis, 2000))
- *width of layout* (the physical width of the drawing should be minimised (Coleman and Stott Parker, 1996))
- *text direction* (all text labels should be horizontal, rather than a mixture of horizontal and vertical) (based on Petre, 1995)
- *font type* (all text fonts should be the same, rather than using different fonts for different types of labels) (based on Petre, 1995)

For UML class diagrams, two additional secondary notation features were investigated. Both versions of each notation have been found in published examples of UML notation (See Figure 1).

- *inheritance* (inheritance lines should be joined prior to reaching the superclass, rather than being represented as separate arcs)
- *directional indicators* (arcs should be labelled with two relationship labels and directional indicators, rather than one)

For UML collaboration diagrams, two additional secondary notation features were investigated. In both cases, the one option (long arrows adjacent to the arcs) is standard UML notation (See Figure 2).

- *adjacent arrows* (all arcs are undirected with an adjacent arrow indicating the direction of the message sent, rather than all arcs being directed)
- *arrow lengths* (the arrows adjacent to the arcs should be the same length as the arcs, rather than shorter than the arcs)

2.3.1 Usability measuring method

As these were our first experiments performed on layout aesthetics and secondary notations for UML diagrams, preference was chosen as the method of

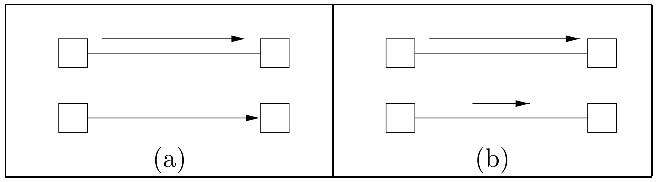

Figure 2: The UML collaboration diagram secondary notation features investigated, showing both alternatives: (a) adjacent arrows, (b) arrow lengths.

usability measurement. While increased preference does not necessarily correspond with improved performance, beginning with a preference study enables the most important layout features to be identified before a more substantial performance study can be performed.

The subjects were chosen as being people who had seen and used diagrams of this nature before. Thus, it is a reasonable assumption that, by asking whether they prefer one UML drawing to another, they are likely to be anticipating the use of these drawings for a software engineering task: their responses are therefore likely to be related to their perceived usefulness of the diagrams.

Three experiments were performed:

Experiment 1 determined the preferences for eight aesthetics embodied in UML class diagrams (70 subjects). Experiment 2 determined preferences for aesthetics embodied in UML collaboration diagrams (90 subjects). In both cases, the quantitative data was the percentage of subjects who preferred one diagram over another, and the qualitative data was the subjects' stated reasons for their choice. The qualitative data was used to determine whether the reasons for preference were linked to the aesthetics under consideration, or whether there were any other unexpected reasons for the choice.

Experiment 3 was a more focussed study which refined the results of the first two experiments, emphasising particular aesthetics that had given surprising results. 6 subjects participated, and extensive qualitative data was collected.

2.4 Methodology

A basic UML class diagram (depicting the relationships between students, lecturers, tutors and administrative staff, see Figure 3) and a basic UML collaboration diagram (depicting the procedure followed for organising honours students' seminars, see Figure 4) were created.

In both cases, the graph structures were complex enough to enable an appropriate and varied manipulation of the nodes and arcs within the diagram, but not so complex that the diagram would take a long time to read.

The class diagram had 14 classes and 18 relationships, and the collaboration diagram had 12 objects and 17 messages.

Each diagram was drawn 16 times (twice for each aesthetic), with all the information within the drawing remaining constant: only the layout of the dia-

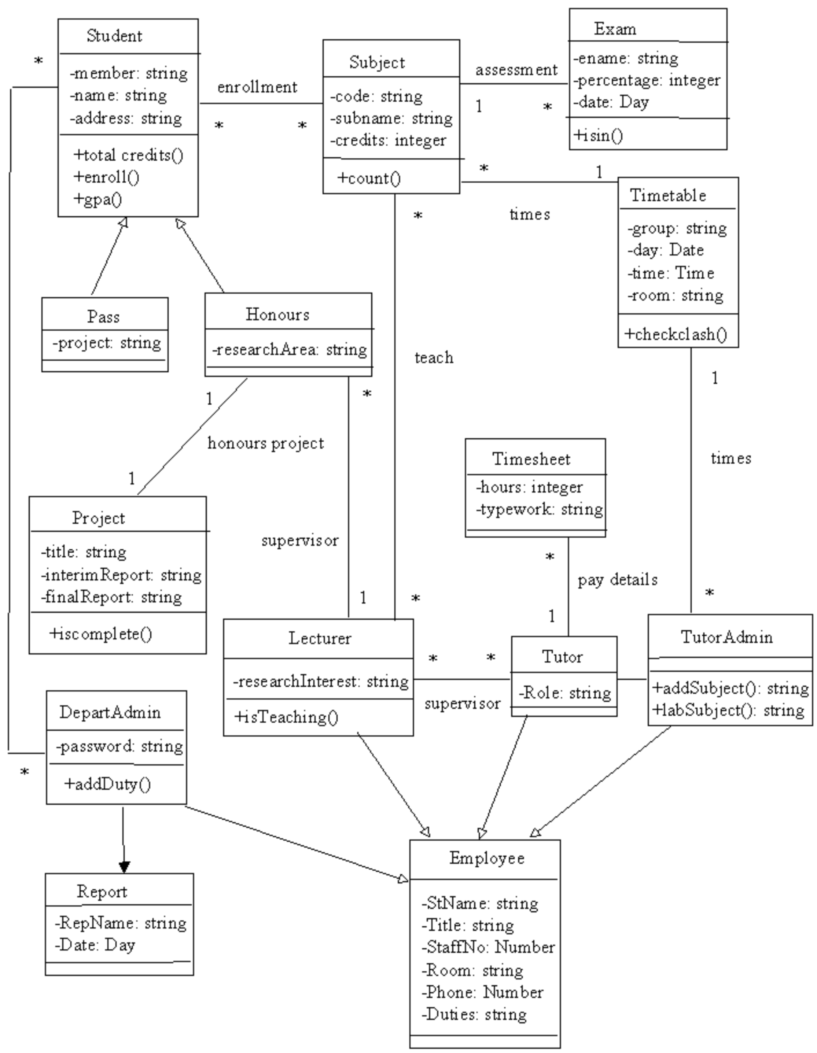

Figure 3: A UML class diagram used in experiment 1.

grams was altered. Each representation of the graph was drawn with attention to a specific graph-drawing aesthetic or secondary notation choice. Figures 9 to 12 shown in the appendix present examples of some of the drawings.

Graph drawings were grouped in pairs emphasising the contrast between the diagrams: one graph drawing in the pairwise comparison contained a higher presence of the aesthetic while the other graph drawing contained a lower pres-

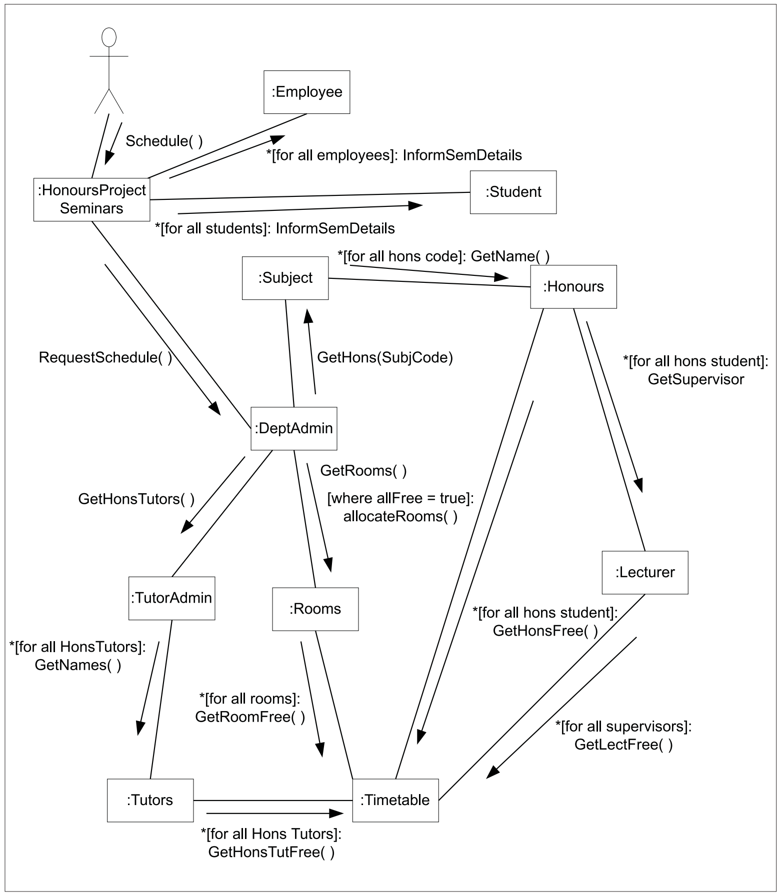

Figure 4: A UML collaboration diagram used in experiment 2.

ence. For example, one diagram was highly orthogonal while the paired diagram had minimal orthogonality.

In the absence of computational metrics for measuring the presence of the UML and secondary notation features in the drawings, and due to the large number of aesthetics being investigated concurrently, the other aesthetics in each contrasting pair could not be controlled. We were aware that not controlling the other aesthetics could have resulted in confounding factors: for example, the diagram with lots of bends may have differed from its paired diagram in both the number of bends as well as in the extent of its perceived orthogonality; the diagram with joined inheritance may have had an increased number of bends. A more robust experiment would have controlled all variables: but this was impossible. Therefore, to prevent our overall conclusions being affected by potential confounding factors which may have biased the simple preference quantitative data, we collected additional qualitative and ranking data. This additional data enabled us to see where subjects may have made a preference choice that was unrelated to the aesthetic being considered, and to consider the quantitative data in the light of these possible confounds. We anticipated that this qualitative data therefore would prove very important in determining our final priority list.

2.4.1 Subjects

Seventy student volunteers from the University of Queensland participated in the class diagram evaluation; ninety students participated in the collaboration diagram evaluation. The same experimental methodology and materials were used for both experiments.

All participants were third or fourth year Information Technology students who, although not generally proficient with UML, have experience with similar notations (e.g., Fusion, dataflow diagrams, entity relationship diagrams and Booch diagrams). The experiments took place during lecture sessions, and participation was voluntary. Some students may have participated in both experiments.

Prior to both experiments, pilot experiments with a small number of subjects were performed to check for problems in experimental materials and procedures.

2.4.2 Materials

Each subject was presented with an individual evaluation booklet designed to be completed without evaluator assistance and without a time limit. The booklet had the following structure:

- Instructions, including the aim of the project, an explanation of the procedure, and an example.
- A questionnaire requesting information about the subject (prior knowledge and use of graph drawings, year of study etc.).
- A UML tutorial sheet, identifying the key points of the diagram under consideration (class or collaboration). This sheet was detached from the booklet so that subjects did not have to turn back to refer to it.
- Eight (facing) pairs of graphs drawings were presented in turn, each pair consisting of a diagram referred to as A, and a diagram referred to as B. At the bottom of the A diagram, the statement "I prefer diagram A to diagram B" was presented (and a comparable, reversed, statement was presented at the bottom of the B diagram). Students were asked to select one of these two statements, and there was a space for them to write a brief explanation for their choice. The explanation was intended to identify any issues influencing the choice that had not been considered, and to provide qualitative data to support the quantitative data. The diagram pairs were presented in a random order in the booklet, in an attempt to counter any

familiarity effect: after seeing several different representations of the same UML diagram, the subject may develop a deeper understanding of the information, and may therefore make different preference choices.

- A ranking sheet where subjects needed to rank the three diagrams that they most preferred (1-3), and the three that they least preferred (14-16). A separate large sheet, showing all 16 (reduced in size) diagrams on one page, was provided to assist in this ranking.

There was no time limit, and the subjects could go back and change any previous answers if they wished. Most subjects completed the task in about 20 minutes.

2.4.3 Data Analysis

The preference questions were analysed by calculating a percentage preference for each aesthetic, with the significance of the result computed using a standard binomial distribution. A result was considered significant (i.e., not attributable to chance or random selections) if its probability was less than 0.05.

The written explanations for preferences were analysed by determining the percentage of subjects who stated that the targeted aesthetic comparison influenced their choice: this allowed us to identify whether there were any possible confounds (i.e., other aesthetics unintentionally affecting the result) in the quantitative data obtained by the subjects' preferences.

The final ranking question was analysed to identify preferred diagrams by computing an overall weighted preference value for each diagram. A weight of 3 was given for a first choice, a weight of 2 for a second choice, a weight of 1 for a 3rd choice, a weight of -1 for a 14th choice, a weight of -2 for a 15th choice and a weight of -3 for a 16th choice.

3 Results

3.1 Results: UML class diagrams

The quantitative results for UML class diagrams are shown in tabular form in Figure 5.

Crosses: 93% of the subjects preferred the diagram with fewer crosses. Of the subjects who chose the diagram with fewer crosses, 88% made direct reference to the number of crossing lines: "Fewer lines crossing", "The other diagram has too many crossing lines." Of the subjects who chose the diagram with many crosses, none indicated that the crosses aesthetic playing any part in their choice; rather, their preference was for the arrangement of the classes "type of students are close, clear, regardless of the crossing lines."

aesthetic choice	% preference
fewer crosses	93%
fewer bends	91%
horizontal labels only	86%
joined inheritance arcs	76%
narrower	73%
more orthogonal	61%
no font variation	61%
directional indicators	60%

Figure 5: Aesthetic preference results for UML class diagrams (all results significant).

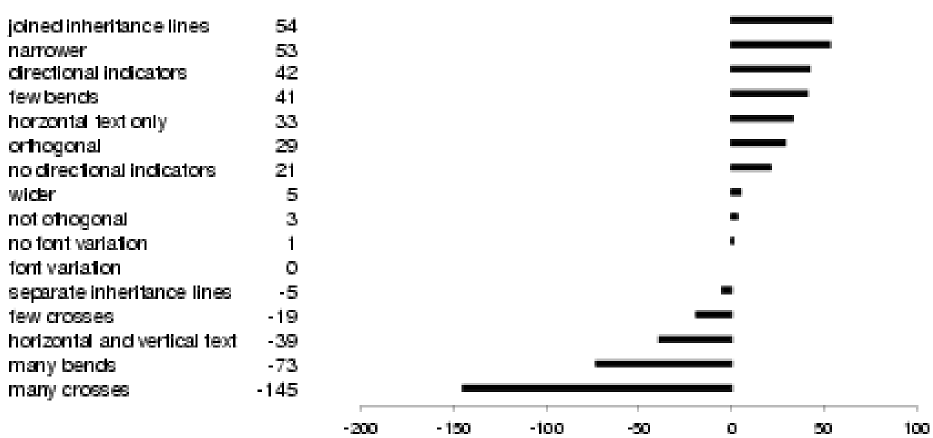

Figure 6: Weighted ranking values for all sixteen class diagrams.

Bends: 91% of the subjects preferred the diagram without bends. Of the subjects who chose the diagram without bends, 52% made direct reference to the bends: "straight lines are better than bent lines"; "Less bends and therefore faster and easier to understand", while 13% commented on the hierarchical layout of this diagram: "Better sorting out of classes (same sort of classes grouped together"). Of the subjects who chose the diagram with bends, none indicated that the bends feature affected their choice: the relationship of classes to one another was a deciding factor ("more common information stored in center").

Text direction: 86% of the subjects preferred horizontal text to a mixture of horizontal and vertical text. Of the subjects who chose the horizontal text diagram, only 20% commented on the text direction: "It is harder to read vertical writing", while 34% commented on inheritance direction "Prefer to follow inheritance from top/left to bottom down." Most of the subjects who chose the combination of vertical and horizontal referred to direction of inheritance flow: "Because most arrows are going down and that is the way we are used to reading things."

Inheritance notation: 76% of the subjects preferred the joined inheritance lines (i.e., as the notation specifies) over the graph drawing method of using separate arcs. Of those who chose the joined inheritance lines, 53% made direct reference to the way in which inheritance was depicted: "shows inheritance more clearly", "more structure." Of the subjects who chose the separate lines, 47% referred to the representation of inheritance: "having individual arrowheads makes it easier to see number of inheritance", "This is more compact."

Overall width: 73% of the subjects preferred a narrower layout over a wide layout. All the subjects who preferred the narrow layout referred directly to the width of the diagram: "the closeness allows you to view more association which helps in following the class diagram"; "[wide] is too hard to follow as the lines are too long." Of those who preferred the wider layout, 90% referred to the width: "[narrow] is too cluttered – it's harder to see the details", "makes it easier to take things in with a quick glance."

Orthogonality: 61% of the subjects preferred the more orthogonal drawing over the non-orthogonal one. Of those who preferred the orthogonal drawing, 62% referred to the underlying grid: "straight and 90 degree angle lines are easier to follow", while an additional 33% referred to structure: "overall layout is more structured." Of those who liked the non-orthogonal drawing, 63% referred to the bends that were introduced in the orthogonal drawing: "lines with bends require more tracing", "straight diagonals are better than L shaped lines."

Font type: 61% of the subjects preferred the use of the same font over varying fonts. Of those who preferred the use of the same font, 63% referred to the use of fonts: "same font allows less distraction"; other subjects referred to the grouping of classes. 65% of those who preferred the use of varying fonts stated this feature as the reason for their choice: "different fonts make it easier to separate sections."

Directional indicators: 60% of the subjects preferred having directional indicators associated with every labelled relationship, rather than not having the directional indicators at all. Of those who liked the directional labels, 88% stated that this was the reason for their choice: "directional labels make arcs more readable", "clearer, precise." Of those who did not like these labels, 75% referred to them : "having two labels on the association lones only complicates it. It's obvious how the relationships work"; "Less complex."

3.2 Discussion: UML class diagrams

By analysing the subjects' stated reasons for their preferences, the only class diagram aesthetic result that appeared to be affected by confounding factors was horizontal labels, where many subjects who preferred the horizontal labels drawing referred to direction of information flow. Both the crosses and bends aesthetic results were unaffected by other factors, while those subjects who did not like the orthogonal diagram did so because of the increased number of bends. Most of those subjects who preferred independent inheritance arcs, a wider diagram, variation in font, or no directional indicators did so merely because of personal preference, rather than because there were other factors that affected their choice.

The stated preference reasons indicated that the direction of flow was an important consideration: this was an aesthetic that had not been included in this initial experiment.

The results of the ranking question (with the ranks scaled with values from 3 to -3) are shown in Figure 6, which presents the weighted overall ranking value for each of the 16 drawings. It shows a distinct dislike for the diagram with many crosses, and preference for joined inheritance lines.

3.3 Results: UML collaboration diagrams

The results for UML collaboration diagrams are shown in tabular form in Figure 7.

Crosses: 91% of the subjects preferred the diagram with no crosses. Of those who preferred the cross-free diagram, 82% referred explicitly to the crosses: "Crossed arcs make it difficult to interpret", "there are no intersections." None of the students who preferred the diagram with crosses did so because of the crosses — direction of flow appeared to be more important: "it filters down"; "I prefer arrows pointing down rather than up."

aesthetic choice	% preference
fewer crosses	91%
no adjacent arrows	90%
longer arrows	82%
no font variation	70%
more orthogonal	63%
narrower	57%
horizontal labels only	54% †
fewer bends	53% †

Figure 7: Aesthetic preference results for UML collaboration diagrams († indicates a non-significant result).

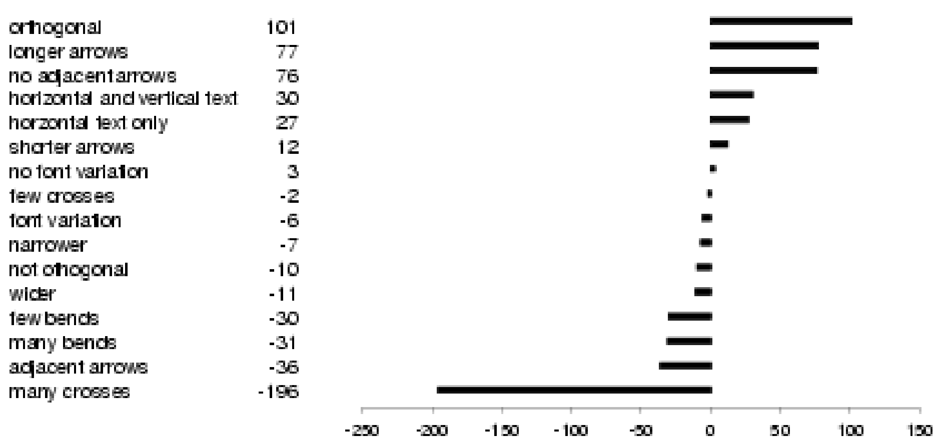

Figure 8: Weighted ranking values for all sixteen collaboration diagrams.

Directional indicators: 90% of the students preferred having a single edge connecting the nodes, rather than an additional directional arc alongside the connecting edge. Of those who favoured not having the additional arrow, 61% made reference to the number of arcs "single connecting line also indicates direction of relationship", "single lines are less confusing, especially at any intersections", "less cluttered." There were no comments on the number of arcs from those subjects who preferred the alternative diagram; 66% referred to direction of flow: "actor at top is more intuitive."

Arrow lengths: 82% of the subjects preferred having longer adjacent arrows. However, of those who preferred the diagram with longer arrows, 57% referred to structure ("neater, better structure", "more ordered"), while only 11% referred explicitly to the arrows: "arrows are longer so easier to see what message is passed." Of those who liked the shorter arrows, 25% referred to the arrows: "shorter arrows still convey the meaning but don't clutter up the diagram", while 19% commented on flow "I like the graph from left to right and top to bottom."

Font type: 70% of the subjects preferred the use of the same font throughout the diagram, rather than different fonts for different label types. Of those who preferred the same fonts, 81% made direct reference to font: "don't like the mixture of fonts/boldness; while 6% referred to crossings "less crossed lines." 62% of the subjects who preferred the different fonts did so because of their personal preference for different fonts: "the change in font makes it easy to see the difference."

Orthogonality: 63% of the subjects preferred the highly orthogonal diagram. 35% of those who preferred the orthogonal diagram referred to the grid structure: "arrows aren't on angles, they are all square"; 51% said they preferred the "more structured" look. Many of those subjects who did not like the orthogonal drawing referred to the increased number of bends (52%) "lines with corners are difficult to follow visually"; "easier to read because you don't have to go around corners when you look at the arrows."

Width of layout: 57% of the subjects preferred a wider layout. Of these, 62% referred to better utilisation of space: "more spread out", while 8% referred to the existence of bends in the alternate diagram: "no jagged lines." Of those who preferred the narrower diagram, 43% referred to the width directly: "it's more compact - easier to take in"; others (24%) referred to the direction of flow: "actor at left, mostly top-to-bottom and left-to-right message passing."

Text direction: 54% of the subjects preferred the use of horizontal labels only (this is a statistically insignificant result). 65% of the subjects who preferred the horizontal text only made reference to the text: 'this diagram does not have vertically positioned text which makes it easier to read". 10% referred to space: "less cluttered." Of the subjects who preferred the diagram with both vertical and horizontal text, none mentioned the text at all: the comments were concerned with structure: "squarer", "structured", "straight lines."

Bends: 52% of the subjects preferred the diagram with few bends (this is a statistically insignificant result). Of those who preferred the diagram with few bends, 38% referred to bends: "prefer straight arrows", while 29% mentioned space considerations"less cluttered and more spread out." The subjects who preferred the diagram with bends made no mention of the bends in the diagram at all. Their comments were related to the single arc cross (24%) "crossing lines is bad", space: "shorter distances between objects" (31%) and flow: "left/right and top/down approach" (19%).

3.4 Discussion: UML collaboration diagrams

By analysing the subjects' stated reasons for their preferences, the only collaboration diagram result that appeared to be affected by confounding factors was longer arrows, where many subjects who preferred the longer arrows referred to diagram structure. Both the crosses and adjacent arrows aesthetic results were unaffected by other factors, while those subjects who did not like the orthogonal diagram did so because of increased bends. Most of those subjects who preferred a wider diagram or variation in font did so merely because of personal preference. The results for the use of horizontal labels only and bends were not significant, so no conclusions can be drawn for these two aesthetics as these results could be attributable to chance.

Like the class diagrams, the effect of flow on the subjects' preferences was evident.

The results of the ranking question (with the ranks scaled with values from 3 to -3) are shown in Figure 8, which presents the weighted overall ranking value for each of the 16 drawings. It shows a distinct dislike for the diagram with many crosses, and preference for orthogonality.

4 Addressing the confounds: the third experiment

It was clear from the comments that the subjects provided as to their preferences that in a few cases, their decision had been made according to factors other than the targeted aesthetic. For example, some subjects preferred the collaboration diagram with both horizontal and vertical text because it appeared more orthogonal.

Experiment 3 was a follow-up experiment that focussed on the aesthetics common to both class and collaboration diagrams, targeting particular aesthetics for which the preference results may have been affected by confounding factors. Its aim was to investigate some of the unexpected results from experiments 1 and 2 that may have been due to confounding factors in the diagrams, and to create a final priority list of aesthetics.

4.1 Direction of flow

The subjects' comments had indicated that the direction of flow of information had sometimes influenced their preference. Direction of flow was an aesthetic feature that had not originally been targeted and which was introduced into experiment 3, with the following definition:

- *flow* (directed arcs should point in a consistent direction, preferably top-to-bottom and left-to-right) (Eades and Sugiyama, 1990)

4.2 The focussed experiments

Six separate smaller experiments comprised this investigation. They were performed in intensive, focussed interviews where each subject was questioned about which aspects of a set of diagrams influenced their preferences. Six subjects took part, each being questioned about all six diagram sets, and providing extensive qualitative data with which to form a prioritised list of aesthetics. None of these subjects had participated in the prior two experiments.

The format of each of the six experiments differed according to the specific investigation of the experiment 1 and 2 results that were being considered. The general procedure was that subjects were shown three or four different diagrams, and were asked questions about their layout preferences.

Most of the diagrams used in experiment 3 were ones that had been used in experiment 1 or 2, although some were altered to target specific aesthetics, or to correct any obvious confounding factors. For example, the original collaboration diagram with few bends had some crossing arcs that had been identified as influencing some subjects' preference for the diagram with more bends and no crossing arcs.

Two categories of unexpected quantitative results were investigated in these six smaller, focussed interview experiments. First, we investigated some surprising overall weighted ranking values. Second, we investigated some aesthetics for which the percentage preferences differed between the class and collaboration diagrams. In each case, we reviewed the diagrams that had been used for experiments 1 and 2 and tried to determine whether there were any confounding factors that may have led to these surprising results.

4.2.1 Overall weighed ranking values:

In both the class and collaboration diagram experiments, the weighted ranking for the diagram with least crosses was negative. It was unlikely that this diagram was ranked according to its lack of crosses: careful inspection of both "least crosses" diagrams revealed that both appeared to be less orthogonal than the other fifteen diagrams in the corresponding set. The first two experiments of experiment 3 focussed on these diagrams.

In experiment 3.1, the original collaboration diagram with no crosses was used, as well as two other diagrams which had no crosses: one which had consistent flow, and one which had high orthogonality. Subjects were asked to rank the three diagrams, and explain their choice. We thought that the original diagram with no crosses had not been favoured by subjects because of its lack of orthogonality, and the outcomes of experiment 3.1 confirmed this: the orthogonal diagram was highly preferred, even though its flow was inconsistent. Some support for flow was demonstrated.

Experiment 3.2 took a similar form, with class diagrams. We thought that the original “least crosses” diagram may not have been favoured because of its single crossed arc and its lack of orthogonality. Two crossless drawings were introduced for comparison and ranking, one which was highly orthogonal, and one which had consistent flow. Like experiment 3.1, the subjects showed preference for orthogonality over consistent flow, and clearly stated that they did not like the single crossed arc.

4.2.2 Difference in percentage preference between class and collaboration diagrams:

Four differences in the preferences for the aesthetics in experiments 1 and 2 were addressed: bends, font variations, text direction and width.

Bends: The results for experiment 2 did not indicate a significant preference for the collaboration diagram with the least bends. Inspection of the pair of collaboration diagrams related to the bends aesthetic revealed that the diagram with more bends had a more consistent direction of flow, and the one with no bends had a single crossed arc. The third experiment of experiment 3 focussed on these diagrams.

Experiment 3.3 used both the original collaboration diagrams with bends, and with no bends. A third diagram was introduced: it was the same as the original diagram with no bends, but with the confound of the crossing arc removed. The fourth diagram used was the same as the original diagram that had bends, but with the bends removed: this diagram had a more consistent flow, and we wanted to determine whether it was because of this flow that the subjects preferred the diagram with no bends.

The results of this experiment supported our view that the direction of flow and the presence of the single cross had influenced subjects’ preference decisions in experiment 2.

Font variation: Using consistent font was preferred more in the collaboration diagram than in the class diagram, and it was felt that this could have been because the collaboration diagram that used a variety of fonts included a more unusual font (cursive) than the corresponding class diagram (italic). This issue was addressed in the fourth experiment by producing identical diagrams with no font variation, and with more subtle font variation (bold).

In experiment 3.4, three collaboration diagrams of identical layout were used: the first used the same font throughout, the second used a cursive font for the object names, and the third used a bold font for the object names.

The results supported our view that the subjects' dislike of font variation is dependent on the type of fonts used, and that the cursive font was particularly disliked by the subjects.

Text direction: The diagram that only used horizontal text was preferred to a much greater extent in class diagrams than in collaboration diagrams. On inspection of the collaboration diagram with both vertical and horizontal text, it appeared to have a more orthogonal shape.

We addressed the possibility of these initial results being confounded by this orthogonal structure in experiment 3.5. In this case, we used exactly the same layout for two diagrams, the only difference being the font direction, where one of the diagrams used both vertical and horizontal text. The third diagram had both horizontal and vertical text, but was more orthogonal.

The fifth experiment confirmed that orthogonality could have affected the subjects' preferences in the first experiment, and was considered as a more important feature than text direction.

Width: There was a stronger preference for narrow width in class diagrams than in collaboration diagrams. Although the qualitative data from experiments 1 and 2 suggested that preference for width may be an inexplicable subjective opinion, we also thought the amount of information associated with arcs in the collaboration diagram may have affected this preference decision.

Experiment 3.6 used two class diagrams with identical object and association layout, one of which was much smaller than the other. The scaling was performed by reducing the arc lengths only: the node sizes remained the same. Two collaboration diagrams with similar size variation were also used. Subjects were asked which they preferred from each pair.

The sixth experiment confirmed that the amount of information on the arcs in collaboration diagrams makes smaller diagrams less attractive. While some subjects could not explain why they preferred the narrower or wider diagram, those that could explain their preference did so by referring to the amount of information in the collaboration diagram.

4.2.3 Conclusions

The follow-up experiments provided rich qualitative data which shed insight on the initial quantitative data. By analysing the extensive interview comments provided by the subjects, we were able to identify which of the aesthetics were most important to them when presented with diagrams that embodied more than one.

We concluded that the priority order of the aesthetics common to both diagram types is: arc crossings, orthogonality, information flow, arc bends, text direction, width of layout and font type. Of the UML-specific aesthetics, we concluded that joined inheritance arcs and directional indicators are preferred for class diagrams. For collaboration diagrams, no adjacent arrows are preferred (although this preference is incompatible with UML notation). This list provides a useful starting point for further studies on UML diagram layout aesthetics with respect to performance in a related task.

5 Discussion

The previous syntactic performance experiment (Purchase, 1997) found support for reducing the number of crossings and bends, and for increasing the symmetry. Information flow and width were not considered.

The results of this semantic preference experiment confirm that the evidence is overwhelmingly in favour of reducing the number of arc crossings as the most important aesthetic to consider.

While the results of the syntactic experiments did not highlight orthogonality as being important, in the domain of UML diagrams, this aesthetic moves up the priority list to second place. In addition, while bends were deemed more important than orthogonality in the syntactic experiments, this was not the case in these UML diagrams. This is an important point when there is an obvious direct relationship between the extent of orthogonality and the number of bends in a diagram.

This difference between the results of the prior syntactic experiments and these semantic experiments is a clear signal that algorithms that are designed for abstract graph structures, without consideration of their ultimate use, will not necessarily produce useful visualisations of semantic information.

This preference study is only the first step in assessing the usability of graph drawings produced by layout algorithms when used in application domains. While preference is a useful start, the real measure of effectiveness of aesthetics is their impact on task performance. Future work includes investigating aesthetics and algorithms with respect to measures of performance in UML related tasks, and extensions of this methodology to other domains.

Graph-based domains like software engineering, social and transport network analysis, and database design are increasingly requiring tools to assist with network visualisation and design. If graph layout algorithms are to be of any use in these areas, it is important that empirical research like that reported here is performed, so that the most appropriate algorithms can be matched with the application domains.

Acknowledgements

We are grateful to the students in the School of Information Technology and Electrical Engineering at The University of Queensland who participated in these experiments. This research was partly funded by an Australian Research Council grant.

References

[1] G. Booch, J. Rumbaugh, and I. Jacobson. *The Unified Modeling Language User Guide*. Addison-Wesley, 1998.

[2] M. Coleman and D. Stott Parker. Aesthetics-based graph layout for human consumption. *Software – Practice and Experience*, 26(12):1415–1438, 1996.

[3] P. Eades and K. Sugiyama. How to draw a directed graph. *Journal of Information Processing*, 13(4):424–437, 1990.

[4] P. Irani and C. Ware. Diagrams based on structural object perception. In *Proceedings of the Conference on Advanced Visual Interfaces, Palermo, Italy*, pages 61–67, 2000.

[5] A. Papakostas and I. Tollis. Efficient orthogonal drawings of high degree graphs. *Algorithmica*, 26(1):100–125, 2000.

[6] M. Petre. Why looking isn't always seeing: Readership skills and graphical programming. *CACM*, 38(6):33–44, June 1995.

[7] H. C. Purchase. Which aesthetic has the greatest effect on human understanding? In G. Di Battista, editor, *Proceedings of Graph Drawing Symposium 1997*, pages 248–259. Springer-Verlag, Rome, Italy, 1997. LNCS, 1353.

[8] H. C. Purchase. Performance of layout algorithms: Comprehension, not computation. *Journal of Visual Languages and Computing*, 9:647–657, 1998.

[9] E. Reingold and J. Tilford. Tidier drawing of trees. *IEEE Transactions on Software Engineering*, SE-7(2):223–228, 1981.

[10] R. Tamassia. On embedding a graph in the grid with the minimum number of bends. *SIAM J. Computing*, 16(3):421–444, 1987.

Appendix

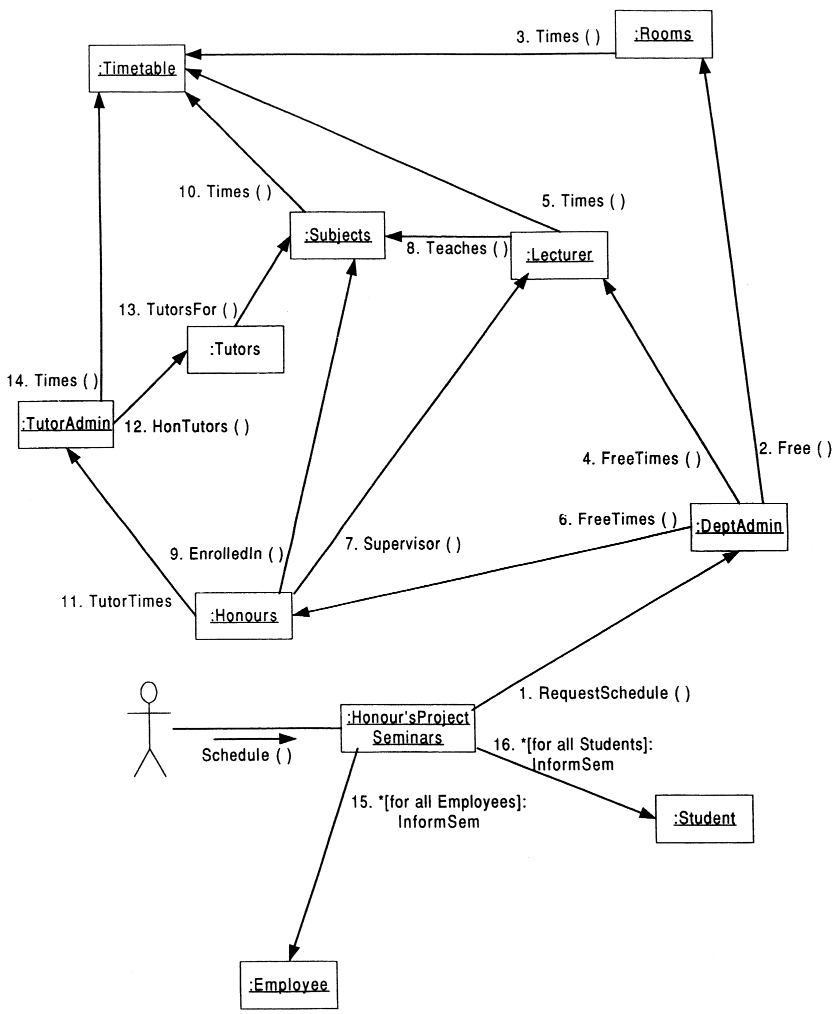

Fig. 9. The collaboration diagram with no arrows.

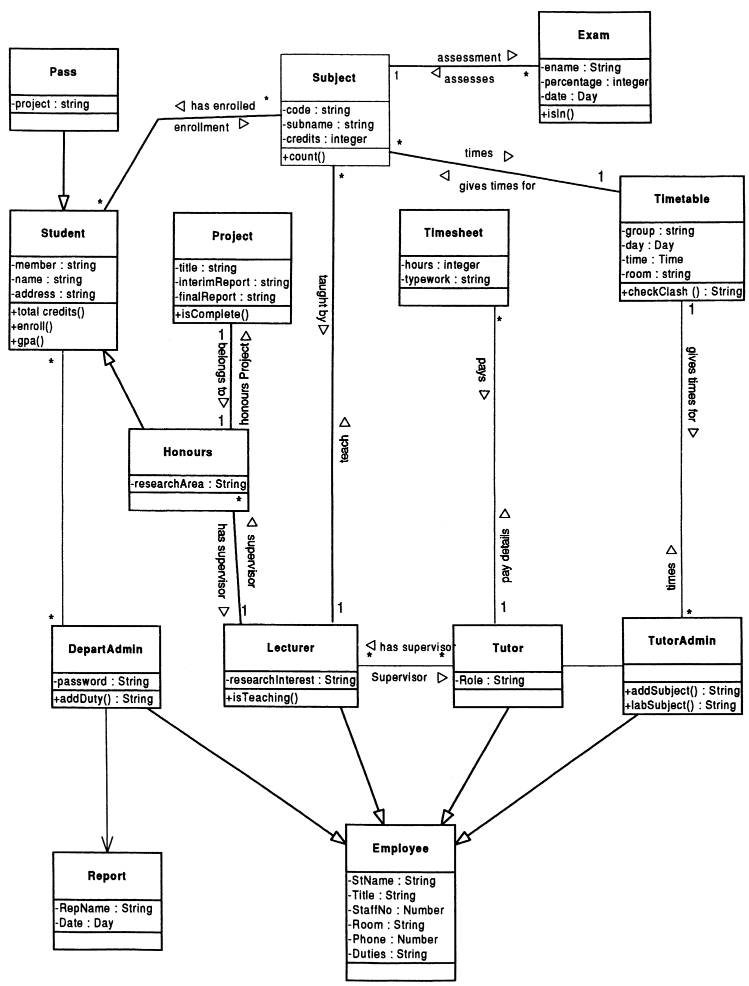

Fig. 10. The class diagram with directional indicators.

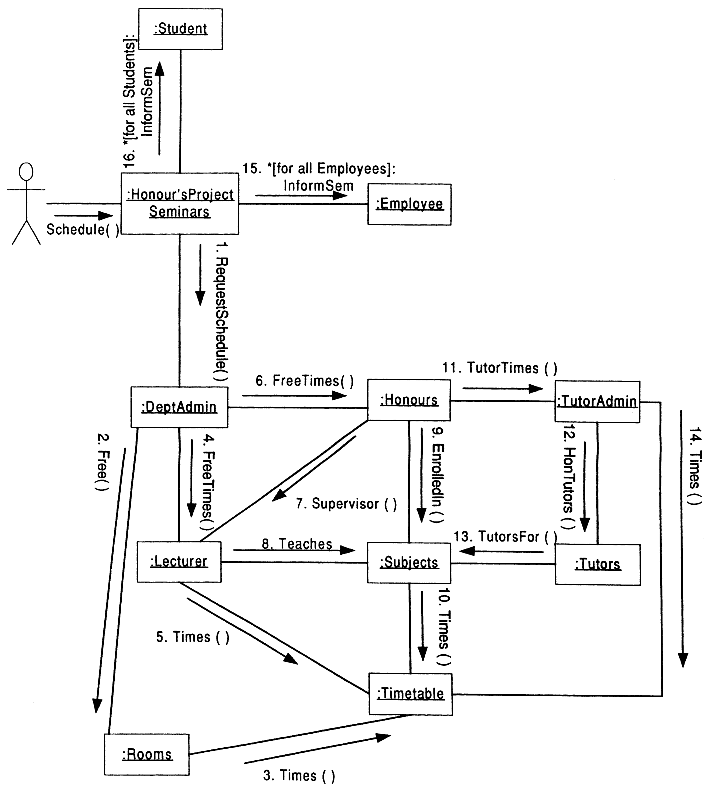

Fig. 11. The collaboration diagram with both horizontal and vertical text.

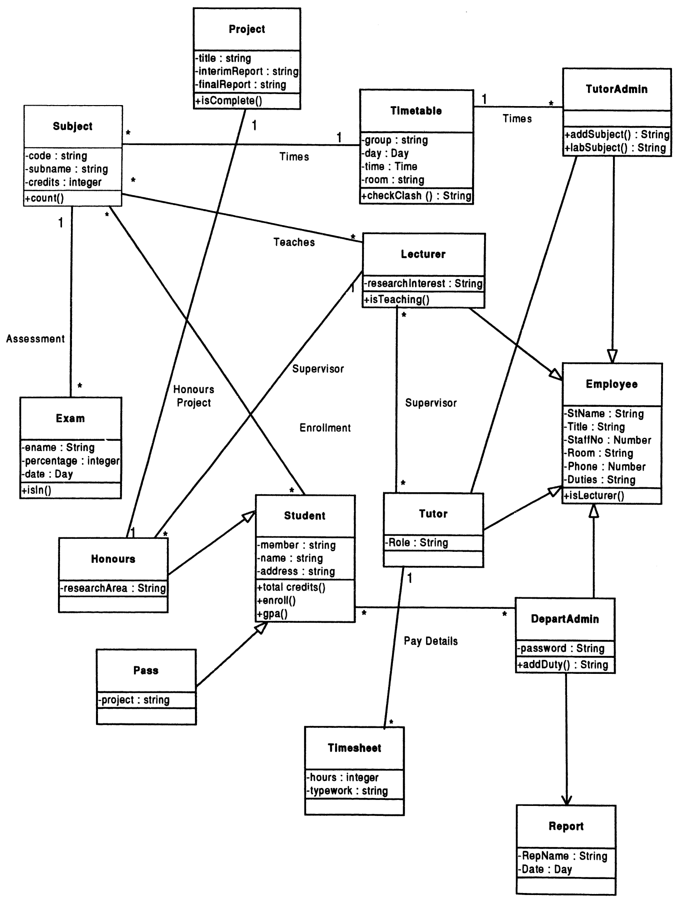

Fig. 12. The class diagram with many crosses.