ANNOTO

AN ANNOTATION FRAMEWORK FOR A COLLABORATIVE LEARNING PLATFORM

SEBASTIAN MADER

Masterarbeit
Betreuer Prof. Dr. François Bry
Niels Heller
Abgabe am 01.03.2016
DECLARATION

Hiermit versichere ich, dass ich diese Masterarbeit selbständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

München, den 01. März 2016

__________________________
Sebastian Mader
ABSTRACT

Annotations play an important role in learning - both in private and collaborative annotation settings. While private annotation and its effects on learning are well-researched, there are fewer studies on digital and collaborative annotation and their effects on learning which leads to the goal of this thesis: Development of an annotation framework which can be used as a foundation for further studies in those areas.

To identify the different annotation types found in digital annotation, an extensive benchmarking of existing annotation platforms was conducted. The benchmarking identified annotation types that are prevalent for annotating a certain type of medium and, therefore, can be considered best practices. Different means for filtering and ordering annotations were introduced to combat the information flood users are confronted with when annotating collaboratively.

The results from the benchmarking were used to create a concept for the annotation framework Annoto which was then implemented in JavaScript. Annoto allows the creation of annotation components which are used for the annotation of a certain type of media. With Annoto media-independent functionality and new types of annotations can be created easily.

Annoto is to be used in a reimplementation of Backstage, a digital backchannel for large class lectures. This reimplementation provides the foundation for further studies in the area of digital annotation in collaborative settings.

Um die verschiedenen Annotationstypen, die beim digitalen Annotieren benutzt werden, zu identifizieren wurde ein umfangreiches Benchmark von bestehenden Annotationsplattformen durchgeführt. Das Benchmark identifizierte Annotationstypen die vorherrschend bei der Annotation eines gewissen Medientyps sind und deswegen als best practices angesehen werden können. Es wurden verschiedene Möglichkeiten zum Filtern und Ordnen von Annotationen eingeführt um das Problem der Informationsflut zu bekämpfen, mit der Benutzer beim gemeinschaftlichen Annotieren konfrontiert werden.


ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Niels Heller. Without his constant support, feedback, and advice the thesis would not be on the current level.

Also, I want to thank Prof. Dr. François Bry for giving me the chance of working on such an interesting topic and offering me a workplace at the research unit.

Thanks to my parents for always supporting me in my studies.

Thanks goes to my closest friends - Christoph, Josip, Marco and Florian - for their support during the last months. Special thanks to Florian for proof-reading the thesis until late at night.

Last, but not least, I would like to thank Iris for always being there for me.
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INTRODUCTION</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>RELATED WORK</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ANNOTATIONS</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.1</td>
<td>Context of an Annotation</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>Referencing, Textual, and Graphical Annotations</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>3.3</td>
<td>Apparent and Non-apparent Annotations</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td>Public and Private Annotations</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>Explicit and Implicit Annotations</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>3.6</td>
<td>Transient and Permanent Annotations</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>Formal and Informal Annotations</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>Lecture and Reading Annotations</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>Different Purposes of Annotations</td>
<td>20</td>
</tr>
<tr>
<td>3.10</td>
<td></td>
<td>Wrapping up Annotations</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>ANNOTATIONS “IN THE WILD”</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.1</td>
<td>Benchmarking Annotation Platforms</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.1 Text Annotation</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.2 Video Annotation</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.3 Image Annotation</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.4 Audio Annotation</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.5 Annotations in Social Networks</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.6 Tagging</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.7 Wrapping up Benchmarking Annotation Platforms</td>
<td>64</td>
</tr>
<tr>
<td>4.2</td>
<td></td>
<td>Creating a Framework for Annotations</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>4.2.1</td>
<td>Classifying the Annotation Platforms</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>4.2.2</td>
<td>Wrapping up Creating a Framework for Annotations</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td></td>
<td>Wrapping up Annotations “in the Wild”</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>COLLABORATING ON ANNOTATIONS</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td></td>
<td>Displaying Annotations</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>5.1.1</td>
<td>Identifying Interesting Passages</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>5.1.2</td>
<td>Wrapping up Displaying Annotations</td>
<td>77</td>
</tr>
<tr>
<td>5.2</td>
<td></td>
<td>Ranking Annotations</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>5.2.1</td>
<td>Creation Time as Ranking</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>5.2.2</td>
<td>Annotator’s Expertise as Ranking</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>5.2.3</td>
<td>Rating as Ranking</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>5.2.4</td>
<td>Filtering Annotations</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>5.2.5</td>
<td>Wrapping up Ordering Annotations</td>
<td>89</td>
</tr>
<tr>
<td>5.3</td>
<td></td>
<td>Wrapping up Collaborating on Annotation</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>BUILDING A FRAMEWORK FOR ANNOTATIONS</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>AnnoTo</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.1.1</td>
<td>Structure of an Annotation Component</td>
<td>92</td>
</tr>
</tbody>
</table>
6.1.2 Annotator ........................................ 92
6.1.3 Media Annotator ............................... 94
6.1.4 Selection ........................................ 95
6.1.5 Annotation Type .............................. 97
6.1.6 Storage ......................................... 98
6.1.7 Content Provider ......................... 98
6.1.8 Wrapping up Annoto ....................... 99

6.2 Annotation Components .................... 99
   6.2.1 Annotation Types .......................... 99
   6.2.2 PDF Annotation ............................ 100
   6.2.3 Image Annotation ......................... 101
   6.2.4 Video and Audio Annotation ............ 102
   6.2.5 Wrapping up Annotation Components ..... 104

6.3 Wrapping up Building the Framework .... 104

7 CONCLUSION AND OUTLOOK .................. 105

BIBLIOGRAPHY ........................................ 109
**INTRODUCTION**

**Motivation.** Annotations are an important part of scientific and academic work. Most students take notes while listening to a lecture, and the majority felt that taking notes plays an integral part in their academic success. [60] A note is defined by the Oxford Advanced Learner’s Dictionary as “information that you write down when somebody is speaking, or when you are reading a book” [31], while an annotation is defined as “note or notes added to a book or text giving explanations or comments.” [31] These two definitions show the difference between the concepts: A note is part of an annotation, namely the content part of an annotation. This difference already provides a simple model of an annotation: An annotation consists of a note, i.e., the content part, and a part of a medium the note refers to, which is in the following called context of an annotation.

First research in the area of annotations and their effect on learner’s performance began early in the twentieth century: At that time, students used notebooks to take notes during lectures. Taking notes was required because books were only available in the library and there was no way to create a copy of single pages. Lecture material was mostly written on a blackboard or mediated through a lecture by the lecturer. Back then, there was no possibility to create printouts of lecture materials, let alone digital versions. Therefore, students’ notes were the main medium for studying for exams, hence, it is not too far-fetched to conclude that students’ academic success was linked to the quality of their notes.

Nowadays, most lecture material is available digitally and one could conclude that this digital availability would lead to a decrease in student annotation. But, when visiting a lecture hall, another image is being presented: Many students take notes, either in a dedicated notebook, in form of annotations on the printouts of the lecture material, or digitally using annotation software. There are two possible explanations for this observation: The first one is that the lecture materials are often only providing an outline, so that students who want to perform well in the exam are forced to expand on the existing notes. The other explanation is that students still see annotation as an integral part of their academic success - and there is a multitude of studies and experiments that back the positive influence of annotation and note-taking on academic success which are further explored in Chapter 2.

The advent of personal computers, laptops, and powerful mobile devices opened up new possibilities and opportunities for annota-
tions. With digital versions of the lecture material available - why even make the detour and create printouts of the material? Technology-enhanced learning tries to incorporate technology into traditional learning with the goal to improve the learning process. [27] Technology-enhanced learning is an umbrella term for a wide field of research and other activities which has produced amongst others Massive Open Online Course platforms, such as edX, Audience Response Systems, such as Backstage or mQlicker, and learning management systems, such as Moodle. As already mentioned, annotations and notes are still an important part in learning and a well-thoughtout annotation component should be part of any collaborative learning platform.

**Contribution.** This thesis reports on the conception and development of the annotation components of a collaborative learning platform that is currently in development. For the implementation of the annotation component, an annotation framework, called Annoto, was developed. With Annoto, so-called annotation components can be implemented which allow the annotation of a certain type of media. Annoto offers basic functionality that is shared between all components without the need to implement that functionality more than once. Extensibility is another feature of the framework: Much consideration was given to allow the simple addition of new annotations and media types. The current implementation does not include all concepts that are discussed in this thesis and has to be seen as proof-of-concept for Annoto and a foundation for further studies in the area of annotations.

The theoretic part explores annotations profoundly and gives an overview of most aspects of annotations. This overview is to give inspiration and a starting point for further development on Annoto and the annotation components. Social aspects of annotation are discussed as well: A problem that arises in a social annotation environment is that there are more annotations compared with a private annotation environment. To combat the annotation flood, users have to be given means for filtering and ordering annotations. For this, different possibilities for determining order are introduced. This thesis explores those possibilities only theoretically to provide a starting point for further development of the framework in this direction.

To identify best practices used for annotation for different media types, an extensive benchmark of existing annotation platforms was carried out. The results were classified using a classification framework developed in concordance with the theoretic foundation. Best practices for each platform were identified and most of them implemented as part of the annotation components.

---

1 [https://www.edx.org](https://www.edx.org)
2 [https://www.mqlicker.com](https://www.mqlicker.com)
3 [https://www.moodle.org](https://www.moodle.org)
Challenges. A big challenge is richness of media: For a collaborative learning platform, different types of media are relevant - there are documents, mostly in form of PDF files, images, videos, and audio files. To some extent, the process of annotating is greatly varying between the different kinds of media. To combat this, the benchmark, as described above, was carried out to identify best practices for each type of media.

The media richness also presented a challenge on the implementation side: The different kinds of media differ greatly in how they are displayed, how their properties are accessed, and how they are modified. Annoto addresses different media types generically so that common functionality is usable across a range of different media types.

Structure. Chapter 2 gives a rough overview of research in the area of notes and annotations and reports on their findings. In Chapter 3, annotations are explored from a theoretical angle. Chapter 3 introduces various classifications for annotations providing the theoretical foundation for the remainder of the thesis. The following Chapter 4 examines annotations from a practical angle: Annotation platforms for different kinds of media are examined and described. Together with the theoretical foundation of the previous chapter, a classification framework for annotation platforms is developed, and the examined platforms are classified. Chapter 5 introduces different possibilities for displaying, ordering and filtering annotations. Different possibilities using data created explicitly and implicitly by the users of the platform are explored. The developed framework is introduced in Chapter 6: Used technology and the structure of the framework are explained in detail. At last, Chapter 7 summarizes the findings and contributions of this thesis and gives outlook on further work in the area of annotations.
This chapter gives a rough overview of research related with notes and annotations both in traditional and digital settings. Most research on traditional note-taking focuses on notes and not annotations. This is due to the time they originate from: Most studies date back to the three first quarters of the twentieth century where it was not possible or common to have printouts of lecture material or textbooks to annotate in the margins. But, as explained in Chapter 1, annotation and note are related concepts, and an annotation can be seen as a note paired with context which makes it possible to apply the results of studies on notes to annotations.

Generally, the presented studies examine annotations from their functions. Di Vesta and Grey [20] attribute two main functions to notes: An external storage and an encoding function. External storage means that notes are providing a place to store knowledge for later review; they offer the possibility to retain knowledge but postpone the transfer of the knowledge to the mind - the internal storage - to a later date. The encoding function means that notes enable the creator to transcribe the knowledge enriched with “subjective associations, inferences, and interpretations” [20, p. 8] which possibly helps with retention and understanding of the material. [20]

TRADITIONAL NOTE-TAKING

The first person to examine the effects of students taking notes in a lecture was Crawford [16]. In a first study, he compared the quality of students’ notes with their test results. The quality of notes was determined by the amount of ideas talked about in the lecture which were also included in the notes and the correctness of the recorded ideas. One result was that “the number of points [meaning ideas, author’s note] recorded in the lecture notes and the number recalled at the time of the quiz” [16, p. 290] are positively correlated. Another finding was that “full, clear, and definitive [notes] are more effective than those [notes] which are brief, sketchy, or vague.” [16, p. 290]

In a follow-up study in the same year, Crawford [15] researched the effect of note-taking in lectures in an experimental setting. In his study, he examined the effects of note-taking in lectures on test performance in three different lecture settings: A traditional lecture, a discussion based on a reading, and a textbook recitation. Two different kinds of tests were examined: A quiz where students had to formulate their own answers to questions and a true-false test where
students had to choose if statements were true or false. In some experiments, tests were written immediately after the lecture, in other experiments there was a set period of time students could use to review their notes before the test was written. [15]

Crawford’s [15] results indicate that students that took notes had an edge in quizzes over students that only listened to a lecture in all cases. While the difference in scores was small when the quizzes were written immediately after the lecture, the difference was significantly greater when students were given time to review their notes. An interesting finding with true-false tests was that when the test was written directly after the lecture, students that only listened to the lecture achieved higher scores than students who took notes. This difference disappeared if the test is taken after a review period, but even then, the test scores of the two groups show a smaller gap than in the quiz setting. Crawford’s explanation for this difference between the two kinds of tests is that true-false tests “measure the ability to recognize specific points, whereas a general quiz measures the ability to recall and organize a series of points.” [15, p. 385]

By showing that the scores of note-takers are higher in the majority of scenarios, Crawford [15] found first evidence for the external storage and encoding function of notes, long before di Vesta and Grey [20] coined the terms in 1972, and created a solid foundation for many studies in the area of note-taking.

Di Vesta and Grey [20] showed in an experiment that taking notes “led to an increase in the number of ideas recalled” [20, p. 13] when students were prompted to write down all the ideas they remembered from a lecture after the lecture. Additionally, taking notes was the only factor in their study that positively influenced results in a multiple choice test. [20]

Fisher and Harris [24] found evidence for both functions with external storage being the function which has more impact on recall of ideas presented in a lecture. In their experiment, 60% of the students were allowed to take notes, while the rest was only listening. The note-taking group was split further into three subgroups: A group who reviewed their own notes, a group that reviewed notes created by the lecturer, and a group that reviewed without notes (called “mental review” by Fisher and Harris). The other group was split into two subgroups: A group that reviewed notes created by the lecturer and a group that did a mental review. This division in five different treatments allowed Fisher and Harris to test both functions of annotations, e.g., comparing the results from the group reviewing their own notes against the group who took notes but reviewed the lecturer’s notes would be a test for the encoding function. [24]

Positive effects on learning from taking notes were found in a study carried out by Carter and van Matre [13]: They found significant evidence for the external storage function of notes but no significant
evidence for the encoding function. According to them, these findings suggest that “it is not note taking, per se, but note having and reviewing which facilitates performance.” [13, p. 903]

In their study, Annis and Davis [3] used both notes created by students and notes created by the lecturer. As with Fisher and Davis [24], this study setup allowed them to measure the two functions of note-taking. Their results show that “both note-taking functions are important, but that the more important function for success in later recall is the encoding function,” [3, p. 46] a result that is contrary to the findings of Carter and van Matre [13] who found no significant evidence for the encoding function. Their conclusion came from the fact that students who took notes and reviewed either their own or notes created by the lecturer achieved higher test scores than students in any other group. For the external memory function, they concluded that the important fact is having the notes, not taking them - reviewing notes, regardless of them being one’s own notes has a positive effect on test scores, which is the same conclusion Carter and van Matre [13] arrived at.

Nonetheless, studies in this area are not conclusive: In another study, Fisher and Harris [25] examined if the form of notes had an effect on recall. In a test written immediately after the lecture, they found no significant difference between students who took no notes, students who only created an outline, and students who took detailed notes. They obtained the same results in a test written after a review period. Their results suggest that note-taking has no influence on academic performance. [25]

For further reading, two excellent overviews of traditional note-taking were written by Carrier and Titus [12] and Mosleh and Baba [53] that, in addition to the aspects discussed in this chapter, cover further aspects of notes and note-taking.

In summary, there is evidence that taking notes has a beneficial effect on students’ performance in tests. While there are studies that found no evidence for this relationship, they are outnumbered by studies showing the opposite. The main issue dividing the studies is which of the two functions of notes the positive influence on students’ performance can be attributed to, with much of the evidence pointing to the external storage function. All of the studies examined traditional note-taking, i.e., using pen and paper. The next section examines the impact of digital note-taking on learning.

**DIGITAL NOTE-TAKING**

Digital note-taking is an area that is not as well researched as traditional note-taking - a fact that is most likely attributable to the comparable small period of time in which solutions for digital note-taking exist.
A pioneer in terms of digital note-taking is Catherine Marshall: To derive implications for digital annotation, she conducted a book store study in which Marshall [45] examined the markings students leave in their college textbooks. According to her, it will be nearly impossible to bring the “endlessly inventive” [45, p. 138] annotations created by students into the digital world without sacrifices. Marshall also argues that the seamlessness of physical annotation is hard to match in a digital annotation environment. Finally, her study introduced design implications for digital annotation software - her most important implications are as follows. [45, p. 139f.]

- **in-situ annotation, distinguishable from the source**: Meaning that annotations can be placed directly on the source material and are distinguishable from the source material, e.g., by a different font style.

- **smooth transitions between private and public annotations**: Annotators should be able to make private annotations public, or vice versa.

- **fluidity in form**: Annotators should not be restricted in their annotation style, e.g., by only offering certain annotation types or colours.

An annotation tool realising all of Marshall’s implications can be seen as the perfect digital annotation tool for readers. But readers - while one of the most important - cannot be seen as the only stakeholders in digital annotation, and the needs of other stakeholders require to find a compromise. For example, for a collaborative annotation platform complete fluidity of form is undesirable because it could potentially lead to annotators creating annotations that only they themselves understand.

A digital note-taking software developed by Davis and Huttenlocher [18] before Marshall’s book store study is CoNote. CoNote is a web application that was used in a computer science lecture at Cornell University. The system was tested under real-life conditions, and the authors admit that most of their findings are anecdotal. They observed that there were fewer students at the tail end of the grade scale which the authors attributed to annotations: Annotations allowed students to ask for help and read what others were asking when working on one of the big four assignments given during the semester. This observation is supported by the fact that more students completed the assignments than in previous years. [18]

Nokelainen et al. [54] used a web annotation platform in a peer based distance learning scenario: Using only the annotation platform, groups of two (randomly assigned) students were tasked to annotate and discuss documents uploaded by the instructor of the course. The same experiment was conducted in a group of post-graduate students.
and a group of master’s students. They found a positive correlation between activity on the system and final grade for both groups. Furthermore, for post-graduate students there was a positive correlation between quality of annotations and final grade. [54]

Hwang et al. [34] introduced a web annotation tool called VPen into a university course to measure the difference in test scores between students using the tool to create annotations during the lecture and students that only read the lecture material. Annotation was examined in three scenarios: Individual annotation, sharing annotations within a group, and sharing annotations among all annotators. Results showed that with one exception students using annotations achieved significantly higher test scores than students that only read the material. In the scenario where annotations were shared among all annotators, the annotators’ mean test score was nearly twice as high as the score of students that just read lecture material. The exception was the last scenario where the test was the final exam which counted towards the students’ final grade in the course. They explained this with the assumption that all students felt compelled to study for the final exam but speculate that students who created and had access to annotations needed less time for preparation. [34]

In a follow-up study, Hwang et al. [33] introduced VPen into a school class of first grade junior high school students. Annotations were used for annotation of learning material but also for creating homework solutions. Students had access to other students’ annotations created both for learning material and homework, creating a kind of peer review environment. They found no correlation between learning achievement and the number of accesses to annotations created by peers. Students mentioned that “it was not easy to get meaning out of the content from text annotations made by others.” [33, p. 1023] The only two factors that had an effect on learning achievement were the number of annotations created by the respective student and the number of homework solutions. Reviewing another student’s homework solutions had no positive effect, but students reviewing other student’s homework annotations admitted that they used annotations created by peers mainly for copying homework and not for finding inspiration and approaches for their own homework solutions. [33]

Su et al. [68] examined collaboration on a web based annotation system in similar scenarios as Hwang et al. [34]: One group had access to a web annotation system called PAMS 2.0, while the other group only read the course material. While Hwang et al. had a scenario where annotations were shared among all annotators, Su et al. only allowed sharing in small groups consisting of two to four students. Findings were similar to the findings of Hwang et al.: In scenarios where test results did not count towards the final grade, students who used annotations achieved significantly higher scores. When test re-
sults counted towards the final grade, results from both groups were nearly identical. Su et al. present the same explanation as Hwang et al., that students felt more compelled to study for exams that count towards the final grade. [68]

In their study, Bui et al. [10] found significant evidence that students annotating a lecture using a word processor are able to record more ideas from a lecture than students who use pen and paper. In their experiment, they divided students into two groups - one annotating using a word processor, the other using pen and paper. Each group was divided further into a group transcribing the lecture and a group organizing the lecture, i.e., adding own ideas and structure to their notes. In further experiments, all students used a word processor for creating notes and were only divided in transcribing and organizing. In tests written immediately after the lecture, the transcribing group achieved significantly better results than the organizing group, while in tests written at a later time without time to review the notes the organizing group achieved significant better results. The results reversed again when both groups where allowed to review their notes before writing the test. To refer back to the functions of a note: Evidence for both functions was found with external storage being the more important function. [10]

While there are studies examining the effect of digital notes and annotation on learning, the area is not as well-researched as traditional note taking. But the direction the existing studies are pointing to is quite clear: Digital annotation, collaboration, and access to other students’ annotations seem to have a positive effect on learning achievement.

WRAPPING UP RELATED WORK

The studies and experiments described above clearly show that in the area of annotations and their effects on learner’s performance there are still aspects where results are inconclusive or which are not researched at all. Most studies examining the influence of notes created by a third person used notes created by the lecturers themselves. With digital annotation, the access to other peoples’ annotations was often restricted to small groups. The difference between lecture and reading notes, introduced by Crawley [17], is an example for an insufficiently examined area: What are the differences between notes created while listening to a lecture and notes created while studying lecture material? Do these differences have any influence in a collaborative learning environment?

Annoto, the annotation framework developed as part of this thesis, has the goal of being a platform that allows the study of annotations. By offering a generic framework, new annotation types and annotation scenarios can easily be created which provide the foundation for
further studies in the area of annotations, especially in a collaborative environment.
Annotations are a wide field of research - while the concept itself may be intuitive and easy to grasp, for the development of an annotation framework that is intuitive to use and easy to grasp, a theoretical and scientific approach is inevitable. As seen in the previous chapter, there is much research in the field of annotations that covers different aspects. The focus of this chapter is form and function of annotations themselves. It is important to examine which different forms annotations can assume: Form does not only include physical manifestation but also other properties, such as comprehensibility and formality. Another aspect, especially relevant in a social annotation setting, is purpose of an annotation: When annotating by oneself purpose and function may be obvious for the creator but probably not for other people.

Different classifications for annotations are introduced in this chapter which are, if possible, supported by studies on the subject. A theoretical approach is important for gaining understanding of the nature of annotations, which is required for the remainder of the thesis.

In the first part of this section, an important term - context of an annotation - is introduced. The latter part of this chapter introduces various classes of annotations.

Classification and concepts found in Sections 3.1, 3.2, 3.3 and 3.7 were already discussed as part of my bachelor’s thesis [43], but are expanded on in this thesis.

3.1 Context of an Annotation

For the remainder of the thesis, the term context will be used to describe the part of a medium an annotation is referring to, i.e., the part of a medium that is annotated. Context is a very important part of an annotation because an annotation - in most cases - refers to a specific part of a medium.

When annotating a passage in a book, different ways for expressing context are imaginable. Two of the possibilities can be seen in Figure 1. The first possibility - on the left side - is by proximity: Context here is expressed by placing a textual comment in proximity to the annotated passage. The second possibility - on the right side - is by using additional mark-up: Context here is expressed by the arrow connecting the textual comment and the annotated passage.

For a collaborative learning platform, not only textual media, but also images, audio, and video have to be considered, therefore, there
are different dimensions for context, which result from the inherent properties of the respective medium.

- **Spatial dimension**: The part of a medium an annotation is referring to can be identified by a spatial property, e.g., coordinates or position relative to the elements of a document.

- **Temporal dimension**: The part of a medium an annotation is referring to can be identified by a temporal property, e.g., a time interval or a point in time.

- **Spatio-temporal dimension**: The part of a medium an annotation is referring to can be identified by a temporal and a spatial property, e.g., a time interval coupled with coordinates.

With physical annotation, users are free in choosing how to express the connection between annotation and annotated part. In contrast, with digital annotation this freedom has to be restricted: The different ways for creating context for annotations have to be implemented in the platform, and therefore, there can be only so much different ways to create context. As a consequence, finding intuitive and easy to use ways to create context of annotations is an important requirement for a digital annotation platform.

### 3.2 Referencing, Textual, and Graphical Annotations

The distinction between referencing, textual, and graphical annotations is introduced by Agosti et al. [2] and is a distinction by the sign - the physical manifestation - of the annotation. This section is based on the findings by Agosti et al. [2].

Examples for the different signs of annotations can be seen in Figure 2. A textual annotation is a textual remark added to the medium or parts of it: In the top left image of Figure 2, the annotation is a textual remark added in the margins of the document. Referencing annotations express some kind of connection between parts of a
Figure 2: Examples for a textual, referencing, graphical, and mixed annotation sign, right to left, top to bottom. Adapted from [2].

When creating an annotation for a physical medium, the possibilities are endlessly and only limited by the imagination of annotators. This is different from digital annotation where annotators are limited by the possibilities offered by the respective platform. Therefore, the distinction between textual, referencing, and graphical annotations is very important because it offers a simple classification for the different signs annotators use to express annotations. This classification is supported by the findings of Marshall’s book store study [45] where the author examined markings left by students in academic textbooks and found markings that all can be classified as either textual, graphical, referencing, or a mix of them.
3.3 APPARENT AND NON-APPARENT ANNOTATIONS

The distinction between apparent and non-apparent annotations is introduced by Marshall [47]. In her work, Marshall uses the term *explicit* for apparent and *implicit* for non-apparent. The change of terminology is done because the two terms implicit and explicit are needed for another distinction in the latter part of this chapter. This section is based on the findings by Marshall. [47]

The difference between an apparent and a non-apparent annotation lies in the comprehensibility of the annotation: While an apparent annotation is understandable for people who are not the original creator of the annotation, a non-apparent annotation may be completely incomprehensible or at least ambiguous for people who are not the author of the annotation.

An example proposed by Marshall for a non-apparent annotation would be an exclamation mark next to a passage of text. For the author of the annotation its meaning is clear, e.g., “this is an important passage”, “this is important for the exam”, or “that can’t be right”, but for somebody who is not the author the meaning remains unclear. On the other hand, a piece of text in the margins of a document is - in most cases - perfectly understandable, even if one is not the author of the annotation.

For a collaborative learning platform, non-apparent annotations are not desirable because if an annotation is only understandable for the author of the annotation, the social aspect of the platform gets lost. Nevertheless, there is an area that uses non-apparent annotations which is especially interesting in the context of a collaborative learning platform: Correction marks used by teachers. If the platform is used for correction, there are generally two ways to solve the problem: Introduce a legend which maps the different correction marks to their meaning or introduce an own annotation type with hardcoded function for each type of correction. In each case, the former non-apparent annotations become apparent. Generally, a social annotation platform should find ways to prevent users from creating non-apparent annotations while restricting their annotation style in the smallest possible way.

An aspect of annotations that is closely related to the distinction between apparent and non-apparent annotations is the distinction between public and private annotations.

3.4 PUBLIC AND PRIVATE ANNOTATIONS

Another distinction introduced by Marshall [47] is the distinction between public and private annotations and is further discussed by Marshall and Brush [48]. As the name suggests, a private annotation is intended only for private use and consumption, while a public anno-
tation is intended for a greater readership. Marshall and Brush conducted a study in which students transferred annotations created on paper into a digital annotation collaboration tool. Their results show that a smaller number of annotations were transferred into the collaboration tool, but those annotations were made more understandable compared to their physical counterparts. Therefore, this distinction also emphasizes the difference in form and comprehensibility when an annotation is created for public rather than private use. [48]

3.5 Explicit and Implicit Annotations

The difference between explicit and implicit annotations lies in the way in which an annotation is created. An explicit annotation is created explicitly, i.e., users consciously interact with the platform with the goal to create an annotation. Implicit annotations are inferred from behaviour on and interaction with a platform and are closely related to implicit feedback which Kelly and Teevan [39] define as “information about users [that are obtained] by watching their natural interactions with the system.” [39, p. 18] For this thesis, an implicit annotation is an annotation created from implicit feedback.

Examples for interactions that can be seen as implicit feedback are the dwelling time on a certain annotation [62, 1], the number of clicks on an annotation [1], or the frequency and number of recurring visits on an annotation. [36] An implicit annotation created from dwelling time would contain user, annotation the user dwelt on, and the dwelling time. Dwelling time and visits yield implicit annotations that allow to draw conclusions about the quality of an annotation: It is likely that an annotation of high quality is visited more often and dwelt on longer than an annotation of low quality. For a collaborative learning platform, implicit annotations are a valuable source of information to calculate quality measures for media and annotations which then can be used to offer means for filtering and ordering to users.

Table 1 shows different behaviours that can potentially be used as implicit feedback and, therefore, to generate implicit annotations. Behaviour category is used to classify the different types of behaviours. For minimum scope the medium interacted with is divided by granularity, e.g., if a certain behaviour is only applicable when interacting with segments of a medium, the scope is segment, or if a behaviour is only applicable to a collection of media, the minimum scope is class.

The behaviour category Annotate is not to be confused with the action of creating annotations considered in this chapter. The behaviour categories are merely used to categorize the different behaviours of users interacting with the medium, so the behaviour category Annotate is the representation of all activities done while annotating an medium that can potentially be used as implicit feedback.
Table 1: Table showing different behaviours that can potentially be used as implicit feedback. Adapted from [39, p. 19] which used [55, p. 484] as foundation.)

Not every behaviour from Table 1 is suitable as implicit feedback on a collaborative learning platform, but the proposed behaviours provide good starting points for further thought about which interactions on a collaborative learning platform can be seen as implicit feedback and as such be used for creating implicit annotations.

3.6 TRANSIENT AND PERMANENT ANNOTATIONS

Transient and permanent annotations are another distinction introduced by Marshall [47]. It is a distinction by the removeability of an annotation from the annotated medium. A transient annotation can be removed from the annotated medium, while a permanent annotation can not be removed and becomes part of the annotated medium. This section is based on the findings by Marshall. [47]

An example for a transient annotation in a physical medium could be a Post-It next to a passage in a book containing the textual remark relevant to the passage. It can easily be removed from the medium. A textual comment written with ballpoint in the margins of a book or highlights done with a text marker can not be removed and, therefore, are permanent annotations.
3.7 FORMAL AND INFORMAL ANNOTATIONS

Another distinction introduced by Marshall [47] is the distinction between formal and informal annotations. This section is based on the findings by Marshall. [47]

A formal annotation is built by rules and follows a syntax which makes the annotation readable - and in certain cases - understandable for machines. An example for a formal annotation would be a key-value pair which is machine-readable and can be made machine-understandable by using a controlled vocabulary for keys and/or values. This concept is, e.g., used for creating meta tags for an HTML document. Another example for a formal annotation are RDF triples where every element of the triple has to be part of a controlled vocabulary.

Informal annotations, on the other hand, are completely free in form, style, and function and follow no fixed syntax. Informal annotations are mostly created by humans using natural language and are intended to be read by other humans. This kind of annotation can not be easily understood by machines, but there are approaches in information extraction that try to extract formal annotations from natural language, mostly by using machine learning algorithms. An overview of information extraction algorithms was created by Wimalasuriya and Dou. [77]

In the context of a collaborative learning platform, both kinds of annotations have their merit: Informal annotations are important for communication and collaboration between users. Formal annotations can be used for other things, such as fuel a search engine which allows to search for media and annotations. Implicitly created annotations are mostly formal because the actual process of creating the annotation is done by a machine.

3.8 LECTURE AND READING ANNOTATIONS

Another distinction is by the situation in which annotations are created: Lecture annotations as opposed to reading annotations. This distinction was introduced by Crawley [17]. Crawley talks about lecture and reading notes, but this distinction can also be applied to annotations because, as already explained in Chapter 2, notes and annotations are closely related concepts. This section is based on Crawley’s findings. [17]

Lecture annotations are created while listening to a lecture as opposed to reading annotations which are created while reading a document. Crawley argues that for lecture annotations to be valuable they have to be revised after the lecture, while revision is not required for reading annotations. The need for revision arises from the circumstances in which lecture annotations are created: Lecturers hold the
lecture, and the annotator has no influence on speed and structure of the lecture, while when creating reading annotations, there is always the chance to turn back a page or read a whole chapter before creating annotations.

3.9 Different Purposes of Annotations

Another important aspect to consider in the context of annotations is the purpose of an annotation. Even if two annotations are represented by the same sign, e.g., a textual remark in the margins of a document, their purposes could be two entirely different things. Examples for the usage of a textual sign include questioning the contents of or paraphrasing a passage.

The purpose of an annotation becomes important as soon as annotations are public because users are not necessarily interested in all kinds of annotations. To distinguish between the different purposes, users would be required to read all annotations, and even then it is imaginable that there are annotations where the real purpose is still not transparent.

A good starting point for devising purposes for a collaborative learning platform are the purposes defined by other annotation frameworks. One of these frameworks is the Open Annotation Data Model developed by the W3C Open Annotation Development Group. The Open Annotation Data Model represents annotations as RDF triples with the goal to create a standard that allows the easy exchange of annotations between different annotation platforms. [28] For expressing the purpose of an annotation, the Open Annotation Data Model defines so-called motivations. The defined motivations are bookmarking, classifying, commenting, describing, editing, highlighting, identifying, linking, moderating, questioning, replying, and tagging. [29]

Another framework for representing annotations in order to make them easily interchangeable is the Annotae project. To express purpose, the Annotae project uses the term annotation class and defines advice, comment, example, explanation, question, and see also as possible values for annotation classes. [38]

The motivations and classes defined by the respective frameworks are in no way exhaustive and most likely not sufficient to cover the different purposes occurring on a collaborative learning platform, but they are a solid starting point. An approach would be to pick suitable purposes from the two frameworks and add further purposes when the need arises.
3.10 WRAPPING UP ANNOTATIONS

This chapter introduced different classes and aspects of annotations. The results of this chapter provide the building blocks for a classification framework for annotation platforms which is developed in the next chapter.

Takeaway from this section is that while physical annotations may seem intuitive and easy to use, for the development of a collaborative learning platform this intuitive knowledge is not enough - a deeper understanding is required which was provided in this chapter.

The next chapter conducts a study of different annotation platforms and uses they classification framework built using the theoretical foundation from this chapter to find the point at which enough platforms were examined.
An important thing to consider when creating a new application in an already established area is that the wheel should not be reinvented. While innovation is important, familiarity with widespread systems or techniques should be preserved. This is no different in the area of annotation platforms. To achieve this, annotation platforms for different kinds of media are examined - benchmarked - in this chapter to get a feel for best practices used in annotating. Another goal of the benchmark is to find inspiration for other features besides creating annotations that are potentially suitable for a collaborative learning platform. The platforms are classified using a classification framework built from a selection of classifications from the previous chapter. The classification framework and the classification of the examined platforms is discussed in the last part of this chapter.

4.1 BENCHMARKING ANNOTATION PLATFORMS

Bhutta and Huq [6] define benchmarking as “the process of identifying the highest standards of excellence for products, services, or processes, and then making the improvements necessary to reach those standards.” [6, p. 254] The benchmarking process is composed of five steps: “Determine what to benchmark, form a benchmarking team, identify benchmarking partners, collect and analyze benchmarking information, and take action.” [6, Figure 2, p. 258] Not all these steps are applicable for the benchmark conducted in this chapter, but they are offering a good foundation for the structured examination of annotation platforms.

This part of the thesis benchmarks annotation platforms for different kinds of media and is interested in the functionality of the platforms. Bhutta and Huq call this a functional benchmarking where the “technology/process in one’s own industry” [6, p. 257] is compared.

Identifying benchmarking partners is the process of identifying the annotation platforms that are benchmarked. Bhutta and Huq explain that the benchmark partners are “companies that are considered by the business community at large to be ‘world class’ in that process” [6, p. 259] As an example how to obtain a list of companies that fit this criterion, Bhutta and Huq mention Balridge Award winners. For annotation platforms, an approach like this is not applicable because there is no authority that rates different annotation platforms. For identifying partners in this thesis a feedback loop was used.
This feedback loop can be seen in Figure 3 which shows the process used to decide if enough platforms for a particular type of media were examined. For a particular type of media, an annotation platform was selected using Google Search or Google Scholar. Following this, the platform was examined: What kinds of annotations are supported? What are the special features of the examined platforms? In a next step, the findings were classified using the classification framework and subsequently used to make a decision if enough platforms were considered for the respective type of media: Are most features of the classification framework covered by the already examined platforms? If yes, consider the next type of media. If no, search another platform for the examined type of media. In the process, the classification framework was extended, when required. The base features of the classification framework are a selection of classifications taken from Chapter 3.

The last step in the benchmarking process - *take action* - is, in case of this thesis, the development of the annotation framework, which is discussed in Chapter 6.

The examined platforms are grouped by media type. If a platform supported more than one type of media, it is listed under the first supported type. The four types are text, image, video, and audio.

### 4.1.1 Text Annotation

Digital text content is ubiquitous - from ebooks through PDF documents to websites. Much of the content on a collaborative learning platform contains large amounts of text, such as lecture slides, tutorial sheets, or scientific publications. A collaborative learning platform without support for text annotation is inconceivable.

An important requirement for being considered as a text annotation platform is the interaction with the annotated text: It must be possible to interact with the text itself, i.e., the platform has to offer the possibility to select parts of text and annotate the selected parts. Platforms where annotations are just anchored at spatial coordinates are therefore not considered for this category.
4.1.1.1 Acrobat Reader

Acrobat Reader\(^1\) is a PDF viewer developed by Adobe Systems. It is available for nearly every kind of device from computers to smartphones. For this chapter, version DC (Document Cloud) on OS X 10.11 was used.

After opening a PDF document, Acrobat Reader offers its users a wide range of annotation possibilities. They can be roughly broken down into three categories: Edit annotations, text annotations, and graphical annotations.

The available edit annotations are shown in Figure 4. They are intended to be used to give feedback to parts of a document, i.e., proposing changes, such as alternative phrasing and adding or removing content. On the top image in Figure 4, the annotator proposes the replacement of *proof of concept prototype* with *application*. The bottom image shows an annotation which proposes the insertion of an additional sentence at the position indicated by the wedge.

![Figure 4: Edit annotations in Acrobat Reader DC: Proposing a change (top image) and proposing to add a sentence (bottom image). Annotated medium: [44].](https://acrobat.adobe.com/us/en/products/pdf-reader.html)

Text annotations available in Acrobat Reader are shown in Figure 5. In the top image, an annotation to a selected passage of text is shown. The bottom image shows a so-called *sticky note* which can be placed at an arbitrary position in the document. The existence of a sticky note is indicated by an icon at the selected position.

The textual comment of edit or text annotations is not always displayed but can be viewed by hovering over or clicking on annotated passages or icons representing a sticky note.

Additionally, Acrobat Reader offers a wide variety of graphical annotations, as shown in Figure 6. The top image in Figure 6 shows passages of text that are highlighted, underlined, and crossed out. *Shapes*, such as the rectangle around the authors in the middle image.

---

of Figure 6, are another type of graphical annotation offered by Acrobat Reader. There are simple shapes, such as rectangles or ellipses, and more complex forms, such as a text box connected with an arrow. Arbitrary-shaped forms, such as the bracket in the bottom image of Figure 6, can be added to the document via a freeform drawing tool.

For most of annotations listed in the previous paragraphs, basic properties, such as opacity, colour, or stroke width can be changed.

Other annotation means include stamps. A stamp is an image placed at an arbitrary position on the document. Examples for stamps that are included by default with Acrobat Reader DC can be seen in Figure 7. Contrary to text, stamps allow annotators to make annotations
more graphic and, therefore, more noticeable, e.g., a green check instead of a textual comment correct.

Figure 7: Examples for stamps in Acrobat Reader DC.

Next to the document a list of all created annotations and their authors is displayed. This overview also shows comments for annotations and allows to comment on annotations. The document - with annotations and comments - can be saved and sent to other people who can add new annotations or comment on existing ones. Collaborative work on a document with live synchronization is not possible.

Acrobat Reader allows the creation of textual and graphical annotations with spatial context. The created annotations are informal and can be made public by sharing a document with created annotations.

4.1.1.2 iBooks

iBooks\(^2\) is Apple’s document and ebook reading application. It allows the import of ebooks and PDF documents. Ebooks can as well be bought via a store that can be accessed using iBooks. The application only allows the annotation of ebooks; PDF documents can not be annotated.

Annotation capabilities of iBooks are basic: Annotators can highlight passages of text in one of five different colours or alternatively underline a passage of text. Optionally, a textual comment can be added to a highlighted or underlined passage. Figure 8 shows how the annotations are displayed in the application: The top image shows a highlighted and an underlined passage. The square next to the highlighted passage indicates that a textual comment exists for that passage. If a user clicks on the square or the highlighted passage, the textual comment associated with the passage is displayed, as can be seen in the bottom image of Figure 8.

Another kind of annotation offered by iBooks are bookmarks: Users can bookmark a page by tapping on an icon representing a bookmark at the top right corner of the page. To indicate a bookmarked page, the icon on the respective page changes its colour from white to red.

iBooks provides an overview of all created annotations and bookmarks: The overview shows the content of the annotated passage and the textual comment if one exists. In case of a bookmark, only the page and the first sentence of the page are displayed. A tap on a bookmark or annotation opens the respective page.

\(^2\) https://www.apple.com/ibooks
One feature worth mentioning is the seamless synchronization of annotations between all devices of a user, i.e., annotations created on one of a user’s devices are immediately available on his other devices.

iBooks only supports private annotations - there is no way to share annotations with the exception of sharing the whole device. The application supports textual and, with highlights and underlining, a subset of graphical annotations. Context of the annotations is spatial, and the created annotations are informal.

4.1.1.3 Google Docs

Google Docs\(^3\) is Google’s web service for creating documents, spreadsheets, and presentation slides from within the user’s web browser. The feature set of those tools is very similar to comparable office applications such as Microsoft Office or LibreOffice.

One of the main features of Google Docs is that it allows the collaborative creation of documents and annotations in real time: Different users can work simultaneously on a single document, spreadsheet, or presentation - added content, changes, and annotations are synchronized seamlessly between collaborators.

Google Docs allows annotation of documents, spreadsheets, and slides. Each of the three types supports different kinds of annotations. Generally, there are three kinds of annotations: Comments, suggestions, and notes which differ in presentation and semantics.

- **Comment**: A comment launches an interaction - other collaborators can add replies to a comment. Comments can be resolved which indicates that comments are intended for discussing some-

\(^3\) https://docs.google.com
thing about a document, e.g., controversial parts. After resolving a comment it is no longer displayed on the document.

- **Suggestion**: A suggestion has the same semantics as an edit annotation in Acrobat Reader. Suggestions can be used to propose different phrasing for a part of the document or the removal of a passage. Suggestions can be accepted whereupon the suggested phrase replaces the former phrase.

- **Note**: A note is a textual comment added to a cell in a spreadsheet.

On a document, collaborators can add comments and suggestions as can be seen in Figure 9. Passages for which a comment exists are highlighted in yellow. A passage can be selected by clicking on it whereupon the passage is highlighted in a darker tone of yellow. The bottom image in Figure 9 shows how suggestions are displayed on the document: The former phrasing is crossed out and prefixed with the suggested phrasing. All comments and suggestions are displayed regardless of being selected in a list to the right of the document. Both images in Figure 9 show either the button for resolving a comment and accepting/declining a suggestion respectively.

The bottom image in Figure 9 shows the buttons for resolving a comment and accepting or declining a suggestion.

![Figure 9: Annotations for documents in Google Docs: A comment for the highlighted passage (top image) and a suggestion proposing to replace the crossed out passage with the underlined passage (bottom image). Both can be resolved or accepted/declined via the buttons shown to the right of the document.](image)

Figure 10 shows annotations on a spreadsheet: The existence of a comment is indicated by a yellow wedge in the top right corner of the cell, and the existence of a note by a black wedge. Clicking or hovering over the cell shows the comment or the note.

Lastly, annotations for presentation slides are shown in Figure 11. Notes referring to the whole slide can be added at the bottom. The
box with the yellow border is a text area for which a comment exists. As with documents, annotations are always displayed to the right of the slide.

Figure 11: Annotations for slides in Google Docs: A comment for a text block (top left part) and its textual content to the right of the slide (top right part), and a comment for the whole slide (bottom part).

Everything created with Google Docs (meaning documents, spreadsheets, and slides) can easily be shared with other users, either by link sharing or by adding individual Google Accounts to the document. Privileges can be set, i.e., if other users are just allowed to view the document, edit the document (which includes creating annotations), or just comment on annotations. It is possible to combine the two sharing possibilities, so that everyone who has the link can view a document, but only a selected set of Google Accounts have the privilege to edit the contents of a document.

Annotation means of Google Docs aim mainly to aid the process of collaborative document creation which is particularly noticeable by
the inclusion of comment and suggestion annotations. Those annotations allow users to discuss questionable parts of a document and resolve comments after a discussion which hides the resolved comment; same with suggestions - they can be discussed and afterwards be either accepted or rejected. While this may be the intended usage of annotations, they can be used in other ways, such as paraphrasing the contents of a passage of text, asking questions about the text, or marking important sections. The prerequisite for this is, however, that every collaborator knows in which way the annotations are going to be used.

Annotations created with Google Docs are either public or private. Supported are mainly textual annotations with a small subset of graphical annotations in form of highlights. Context is spatial and the annotations informal.

4.1.1.4 Kindle

The Kindle started as Amazon's ebook reader but evolved from being just an ebook reader with an integrated store to a whole ecosystem spanning different devices, applications, and services. For the ebook reader, there are different product series, such as the Kindle Paperwhite or the Kindle Fire. There are smartphone applications that allow users to read their purchased ebooks as well.

With small differences, the reading applications and the ebook reader support the same kind of annotations. This section focuses on annotations created on a Kindle Paperwhite. Readers can highlight passages of text, and optionally add a textual comment. This can be seen in Figure 12: The right picture shows a highlight, while the left picture shows a highlight with a textual comment. The existence of a textual comment for a highlight is indicated by a small number at the end of the highlight, as can be seen in the left picture in Figure 12.

Readers can create bookmarks for positions in books which allows them to easily switch between different parts of a book.

So far, Amazon's approach to annotation is pretty similar to Apple's with iBooks, but the two platforms differ in one important aspect: Amazon allows sharing of annotations, while Apple does not. Amazon supports two kinds of sharing: Implicit and explicit sharing.

All highlights created by readers are implicitly aggregated by Amazon to identify passages of text that a certain percentage of readers have annotated. These text passages are displayed underlined with a remark how many users annotated that passage. The left picture in Figure 13 shows a passage that was highlighted by 234 readers. From experience, those passages are mostly prominent quotes, e.g., “One Ring to rule them all (…)” from Tolkien’s Lord of the Rings.

Additionally, readers can share their annotations explicitly by making all of their annotations for a book public which can be done via
Figure 12: Display of private annotations on Amazon’s Kindle: A comment for the highlighted passage (left image) and a highlighted passage without comment (right image). Annotated medium: [50].

Figure 13: Display of public annotations on Amazon’s Kindle: An implicitly aggregated highlight (left image) and a public comment from a subscribed user (right image). Annotated medium: [50].

the Kindle website. Users can subscribe other users whereupon public annotations of the subscribed users are displayed in their own books, as can be seen in the right image in Figure 13. The image shows the annotation of the user Dexter Deadwood for the underlined passage.

With Kindle, annotations are private but can be made public. The supported types of annotations are mainly textual, but with highlights a small subset of graphical annotations is supported. Context is spatial because annotations can be created for arbitrary passages of text. The created annotations are informal.

4.1.1.5 AnnotateIt

AnnotateIt is a web platform that allows annotation of websites. For the annotation process, a bookmarklet (a piece of JavaScript code dis-
guised as a bookmark that is executed when the bookmark is called) is used. To create annotations for a website, users simply visit the desired website and call the bookmarklet.

![Figure 14: Display and creation of annotations with AnnotateIt: Dialogue window to create an annotation for the selected passage of text (left image) and the dialogue window shown when hovering over an existing annotation (right image). Annotated medium: PMS website.](image)

After calling the bookmarklet, users can start to create annotations. To create an annotation, users select arbitrary parts of text on the website. Once the selection is finished, a dialogue window is shown which allows the user to add a textual comment and tags. The dialogue window can be seen in the left image of Figure 14. The two checkboxes allow to set if the annotation can be viewed and/or edited by other users. The existence of annotations is indicated by a yellow highlight of the annotated passages, as can be seen in the right image of Figure 14. Hovering over the highlighted part displays the annotation and tags.

Annotations that have been made viewable for everyone are displayed for other users which annotate the page the annotation is referring to. Therefore, the privacy of an annotation can be controlled on annotation level.

AnnotateIt allows either public or private annotations. Context is spatial, and the annotations are informal. The created annotations are mainly textual, but with highlighting passages of text a small subset of graphical annotations is supported.

### 4.1.1.6 Annotate.co

Annotate.co is a commercial web platform for collaborative annotation of documents and images. To start creating annotations, users first have to create a workspace and (optionally) add other people (called collaborators) to the workspace. In the free version - which is considered here - collaborators can be either viewers or users. The difference between the two roles is that while viewers are only allowed to view documents and annotations, users can add new documents and annotations.

6 http://www.en.pms.ifi.lmu.de/research/index.html
7 https://www.annotate.co
For text annotation, Annotate.co supports the basic annotations already seen on the various platforms examined before: Passages of text can be highlighted, crossed out, or underlined and optionally be accompanied by a textual comment. Edit annotations as with Google Docs and Acrobat Reader are supported as well. Annotate.co supports handwritten annotations which are textual annotations referring to arbitrary-shaped polylines on the document. Figure 15 shows the display of annotations on the document: The left image shows a highlight and the associated textual annotation. On the right picture a handwritten annotation, and the associated textual comment is shown.

![Figure 15](image)

Figure 15: Annotations for a text document on Annotate.co: A textual comment for the highlighted passage of text (left image) and a textual comment for the polyline on the document (right image). Annotated Medium: [44].

When annotating images, annotators have different possibilities to define the part of the document an annotation is referring to: By default, rectangular areas are created, but it is possible to create arbitrary-shaped polylines. These two possibilities can be seen in Figure 16: An arbitrary-shaped polyline in the left image and a rectangular region in the image on the right.

![Figure 16](image)

Figure 16: Annotations for an image on Annotate.co: A polyline with a comment (left image) and a rectangular area with a comment (right image). Annotated medium: US Forest Service photo by Susan Blake.⁸

Annotate.co offers the possibility to create referencing annotations: Every annotation can additionally to its proper sense, e.g., being a textual comment, be used as a reference to another part in the same or another document in the workspace. The reference leads to the refer-

⁸ Available online: https://www.flickr.com/photos/forest.service.southern_region/18929553608
enced part if clicked on. Every kind of annotation can be commented on.

Tags are handled differently by Annotate.co compared to other annotation platforms: Annotate.co makes a distinction between shared and private tags. While shared tags are visible for every collaborator, private tags are only visible for the creator of a tag. Such a distinction is useful in workspaces with a greater number of collaborators where public tags potentially come from a controlled vocabulary - here, single users can still attach their own tags to annotations without clashing with the controlled vocabulary. As opposed to other platforms, tags are not added by typing but by selecting the tags from a list which makes it easier to follow a controlled vocabulary.

Annotations are not synchronized seamlessly: While it is possible for two or more collaborators to create annotations for a document at the same time, annotations are not automatically synchronized between collaborators. The document has to be reloaded before collaborators are able to see the annotations created by others.

The visibility of annotations can be set on annotation level: An annotation can either be private, i.e., only visible to the creator, visible to all collaborators, or a set of collaborators. Annotate.co allows the chronological filtering of annotations which, e.g., can be used to only display annotations of a certain age. The display style of annotations can be changed as well: Annotations are either displayed directly on the document, to the right of the document, or below a page in the form of footnotes.

Context supported by Annotate.co is spatial and every kind of sign - textual, referencing, and graphical, is supported. The created annotations are informal.

4.1.1.7 Conclusions for Text Annotation

The previous section introduced different platforms and applications for annotating text documents. There are features that are supported by nearly every platform but also features that are unique to a single platform. In this section, features and approaches to text annotation that are important and suitable for a collaborative learning platform are identified.

Textual content on a collaborative learning platform includes lecture slides, lecture notes, tutorial sheets, and scientific papers. The core annotation features for such content are highlighting, underlining, and crossing out passages of text and optionally adding a textual comment. This conclusion is backed by the fact that every examined platform supported some kind of highlighting.

Google Docs, Annotate.co, and Acrobat Reader offer annotators the possibility to create edit annotations, i.e., annotations that propose removal or alternative phrasing of passages of text. Edit annotations are required as soon as platforms aim to offer the collaborative cre-
ation of textual content, which is obviously a goal of a collaborative learning platform, e.g., the collaborative creation of a seminar paper or presentation.

Referencing annotations are required as soon as links within or between different documents support the understanding of the content. On a collaborative learning platform such links are important. An example would be a lecture where referencing annotations are used to link occurrences of terms to slides that explain the term.

Annotation in form of freeform drawing is a double-edged sword: On the one hand, it greatly improves the expressivity for users - users can use the kind of annotations and signs they are familiar with - but, on the other hand, it is conceivable that freeform annotations easily become non-apparent. Non-apparent annotations are not an issue in a private annotation environment, but on a collaborative learning platform, annotations should be exclusively apparent. Nevertheless, for certain topics freeform drawings are vital, e.g., in a maths lecture, users should be able to easily write down mathematical formulas without using complicated markup.

Tagging of annotations is reasonable to make annotations searchable without only resorting to the content of an annotation. A distinction between private and public tags, as seen with Annotate.co, can be reasonable and improve collaborative working with documents but also introduces more management overhead and could potentially deter users from using tags. Tags are explored in a later part of this chapter in detail.

Additionally, users should be able to place annotations everywhere on the document, as with Acrobat Reader’s sticky notes. This is important because annotations are not necessarily related with a passage of text or span a greater passage of text where highlighting the whole passage would be detrimental, e.g., when writing a summary of a whole chapter or paragraph.

On a collaborative learning platform, commenting on annotations should be possible to allow discussion or the correction of content, e.g., if a learner created an annotation that includes a misconception.

Textual content is one kind of content that plays an important part for a collaborative learning platform, but there are also other important kinds of media, such as video content.

4.1.2 Video Annotation

A collaborative learning platform without support for annotating videos would feel incomplete - there is a great variety of learning content available in form of videos: There are lecture recordings, videos explaining certain topics, or video lectures created exclusively for online consumption. Hence, it is important to provide users with the
tools to work with video content - and one important aspect of working with video content are annotations.

This section introduces different platforms that allow the annotation of video content.

4.1.2.1 YouTube

YouTube\textsuperscript{9}, the world’s largest platform for uploading and distributing videos, offers creators of videos the possibility to add annotations to their own videos. Only the creator of a video can create annotations; other users have no option to create annotations.

YouTube supports five kinds of annotations that mainly differ in how they are displayed on the video. Each kind of annotation is basically a rectangle which dimensions can be changed to match the annotator’s requirements. A time interval is associated with every annotation - annotations are displayed on the video for the duration of the time interval. Basic properties, such as border colour, background colour, and text size, can be adjusted if applicable for the respective annotation. Figure 17 shows the display forms of the five annotation types.

![Annotations available on YouTube: Note, speech bubble and title annotation (top images), a spotlight annotation in normal (middle left image) and hovered state (middle right image), and a label annotation in normal (bottom left) and hovered (bottom right) state. Annotated medium: Birds and Feeder by Jeffrey Beach.\textsuperscript{10}](image)

A speech bubble, which can be seen in the top middle image of Figure 17, is a rectangle with a wedge on the bottom mimicking a speech

\textsuperscript{9} https://www.youtube.com

\textsuperscript{10} Published under CC BY 3.0 (http://creativecommons.org/licenses/by/3.0). Changes from source material: Annotations. Available online: https://archive.org/details/BirdsAndFeederroyaltyFreeStockVideo)
bubble. Annotators can add text which is displayed within the rectangle. The top left image of Figure 17 shows a note which is basically a speech bubble without the wedge on the bottom of the rectangle. The title annotation, which can be seen in the top right image of Figure 17, is a transparent, borderless rectangle in which text can be inserted.

Spotlight and label annotations add additional content to the annotation if they are hovered by users. A spotlight annotation consists of two opaque rectangles which can be positioned independently on the video. One of the rectangles has a border and contains no text, while the second one contains text. The second rectangle is hidden until a user hovers the first rectangle. The two images in the middle of Figure 17 illustrate this behaviour: The first picture shows the non-hovered state, the second one the hovered state.

Label annotations are similar to spotlight annotations but only consist of a single rectangle, and the text that is displayed when hovered is always positioned at the bottom of the rectangle. An example for label annotations can be seen in the bottom images of Figure 17.

YouTube allows annotations to act as a link but severely restricts the destinations links are allowed to point to: Links can only point to locations within YouTube, such as another video, a Google+ page, or a fundraising project. Accounts can be approved for external annotation links which allows to link to a number of approved retailers.11

Annotations on YouTube are inherently public, and there is no possibility to create private annotations with the exception of setting the visibility of the whole video to private. As already mentioned, it is not possible to create annotations for videos of other users. This suggests that annotations on YouTube are mainly intended to be a medium for creators of videos to communicate with their viewership. These annotations are not open for discussion and are, according to experience, often used to correct mistakes in videos, link to subsequent videos, or as a subscription link for the channel of the creator.

Context of every type of annotation is spatio-temporal. Annotations are of a textual nature and informal. The length of textual content is limited by the dimensions of the video because every annotation is displayed directly on the video.

4.1.2.2 VideoANT

VideoANT12 is a video annotation tool developed by the University of Minnesota. It was developed to support video-based peer review: In a video-based peer review, recorded student presentations are made available through the tool for annotation by other students with the goal to provide feedback on the presentation. Before development on the tool started, Hosack [32] conducted a small survey of similar

11 https://support.google.com/youtube/answer/2760471?hl=en
12 https://ant.umn.edu, new version, the version used in this thesis no longer available
video annotation platforms. The examined platforms were measured against three main requirements for the project: Support for large amounts of text, synchronization of video and annotations, and ease of use. [32]

Users have two possibilities to add a video for annotation: A video can be directly uploaded via an external URL or taken from YouTube. After adding a video, users are presented with the interface that can be seen in Figure 18. The play button starts playback of the video, and users can add an annotation for the current position of the video by clicking the Add an annotation button. Annotations are created for a point in time as opposed to an interval as with YouTube. With VideoANT, annotations consist of a subject and an optional textual comment. Created annotations are displayed ordered by timestamp on the right side of the video, as can be seen in the top image of Figure 18.

Figure 18: Annotations with VideoANT: Three annotations and their representation on the scrub bar and the list to right of the video (top image) and blue highlighting of an annotation which was created for the current playback progress (bottom image). Annotated medium: Sunrise Time Lapse by Jeffrey Beach.\footnote{Published under CC BY 3.0 (http://creativecommons.org/licenses/by/3.0). No changes from source material. Available online: https://archive.org/details/SunriseTimeLapse}
When videos are played back, and the timestamp associated with an annotation matches the current position of the video, the annotation is highlighted in the list to the right of the video. In the example, the annotation highlighted in blue is currently active. This can be seen in the bottom image of Figure 18. Hosack [32] states that the choice to display textual comments not on the video was done to prevent cluttering and allow long, textual annotations.

Annotations can be shared via link-sharing: Users that have the link can create annotations or comment on existing annotations - hence, annotations can be either public or private. Because annotations cannot be placed on the video and are only associated with a point in time context is temporal. Annotations are limited to textual annotations and informal.

4.1.2.3 ANVIL

ANVIL[14] is a video annotation tool developed by Michael Kipp, a professor at the University of Applied Sciences in Augsburg. The tool is mainly intended for the formal annotation of videos for scientific use, such as in linguistics. A more thorough overview of ANVIL and its capabilities was written by Michael Kipp. [40]

In ANVIL an annotation (called annotation element) belongs to a track. Tracks are user-defined and can be organized into track groups. A track can be used to group together annotation elements that are semantically related. Another example, mentioned by Kipp [40], would be a track group for gestures containing a track for every person in the video. The top image in Figure 19 shows the scrub bar with a track group called Test group containing two tracks called First test track and Second test track. The second track contains an annotation element indicated by the differently shaded bar. The width of the bar indicates the time interval the annotation is associated with.

Groups and tracks can be specified in a so-called coding scheme which is independent of annotated videos, i.e., a coding scheme can easily be reused for other videos. While ANVIL supports informal annotations in form of a comment associated with an annotation element, its major strength lies in the creation of formal annotations. For tracks and groups, attributes with a controlled vocabulary as possible values can be defined. After that, when creating an annotation for a certain track, annotators have to enter values for every attribute while conforming to the specification of the coding scheme. There are simple values, such as string, boolean, or numbers from a user-defined range, but also more complex attributes, such as a reference to another annotation element.

The image in the middle of Figure 19 shows the creation of an annotation item for a track. The dialogue window in the right part of

---

14 http://www.anvil-software.org
Figure 19: Annotations with ANVIL: The scrub bar containing an annotation item in Second test track (top image), the dialogue window shown after defining an interval and track for creating an annotation item (middle image), and the display of the content of an annotation item (bottom image). Annotated medium: Test video for ANVIL.\textsuperscript{15}

Figure 19 is shown after an annotator chose a track and an interval by dragging an interval in the respective part of the scrub bar representing the desired track. The dialogue window indicates that for an annotation element on this track a number attribute in the range from 0 to 9 and a point attribute is required. The point attribute consists of a number of coordinates on the video which can optionally be connected. This can be used to display polylines on the video. The bottom image of Figure 19 shows the display of an annotation after the annotation item was selected from the scrub bar.

ANVIL’s powerful specification editor (which is the tool used to create a coding scheme) allows the definition of every kind of context. By default, ANVIL only supports temporal context, but by adding a

\footnote{Available online: http://anvil-software.org/download/index.html}
point attribute to a track or group an annotation element can carry spatial information too. The specification editor allows the creation of textual and referencing annotations. The created annotations are mainly formal because they have to follow the coding scheme for the respective track or group. The annotations can be shared by exchanging the file containing the annotations.

4.1.2.4 **Mediathread**

Mediathread\(^{16}\) is an annotation platform that “supports multimedia analysis within a communal environment” \([70]\) and is developed by the Columbia Center for New Media Teaching and Learning. Mediathread allows the collaborative annotation of video, audio, and image content.

The platform is mainly intended for usage within a university course: An instructor posts assignments and defines online collections from where students are allowed to take material from. An example for an online collection is Wikimedia Commons, for material an image from Wikimedia Commons. Afterwards, students of the course can add material from an online collection to Mediathread via a bookmarklet. Added material can be annotated by every user of the course. The material and the associated annotations are then used to compose essays. Mediathread has been used in over 250 courses at the Columbia University. \([70]\)

Regardless of the type of the annotated medium, every annotation is required to have a title and optionally a textual comment and tags.

Annotations for a video in Mediathread are associated with a start and an end time, i.e., video annotations are always created for a time interval. Figure 20 shows the interface of video annotation with Mediathread: Annotations for a video are displayed to the right of the the video. Annotations can be ordered either by author, i.e., all annotations from an author are grouped together or, as in Figure 20, by tags. Ordered by tags, annotations with the same tag are grouped together which means that annotations with more than one tag appear in more than one group. Every group gets a random colour assigned.

The two bars below the scrub bar of the video, which can be seen in Figure 20, are used to visualize annotations. The first bar is used to visualize the time interval of the currently selected annotation: In this case the time interval from 2:54 to 3:11 for the annotation Test annotation. The second bar is used to visualize the time intervals of all annotations for the video. For this, the randomly assigned colour of the respective group is used.

Audio annotation is basically the same as video annotation which is attributable to the fact that Mediathread only allows temporal context for video content.

\(^{16}\) http://mediathread.info
When annotating images, users first have to choose the part of the image they want to annotate: An annotation can either be made for a single point, i.e., a spatial coordinate, or an user-defined closed polygonal area, i.e., a set of spatial coordinates connected by a line. Figure 21 shows two polygonal regions: One around the soldiers in the bottom half of the image and another one around the tree in the top half. The cannon in the top right half of the image is selected with a point selection. When opening an image, no annotation is displayed on the image, but a list of all annotations is shown to the right of the image. When clicking on an annotation in this list, the region of the annotation is displayed on the image. Users have the possibility to display all annotations which displays the different selections in the assigned colours.

The main reason for annotating in Mediathread is to use annotations to create compositions. A composition is a combination of written text and annotations together with the part of the medium they are referring to. The resulting composition can be viewed together with the annotations - the annotations are clickable and a click shows the annotated medium and the selection associated with the annotation, e.g., clicking on an annotation associated with a video displays

Available online: https://www.youtube.com/watch?v=d0WfO4drHHA
Figure 21: Image annotations on Mediathread: The right parts shows the image and the regions for which annotations exist. The left part shows the textual comments ordered by tag. Annotated medium: The 9th New York Infantry Regiment charging the Confederate right at Antietam by U.S. Army.

the video and starts playback of the time interval associated with the annotation.

Mediathread’s focus lies on collaborative work and collaborative creation of compositions, and therefore, there is no possibility to create private annotations. Context for audio and video annotations is temporal, for image annotation spatial. Mediathread only supports informal, textual annotations.

4.1.2.5 Conclusions for Video Annotation

The previous section introduced different platforms that used different approaches for video annotation. Again, same as with text annotation, there is no silver bullet for video annotation - different kinds of platforms have different sets of requirements. This section considers the different annotation features for videos in the context of a collaborative learning platform.

One of the main distinctions of the different approaches is context: Is spatio-temporal context required or is temporal context sufficient? Support of spatio-temporal context is indicated if the annotated videos contain a great amount of spatial information, e.g., a video of a lecture where the lecturer makes extensive use of the whiteboard as opposed to a video of a lecture which only shows the head of the lecturer. For a collaborative learning platform, both kinds of videos are conceivable, therefore, spatio-temporal context should be supported.

Spatio-temporal context introduces further challenges for displaying annotations because in this case there is additional spatial infor-
4.1 BENCHMARKING ANNOTATION PLATFORMS

Annotation associated with annotations that has to be displayed. YouTube displays annotations and the associated text directly on the medium. This approach has drawbacks: The character count of comments is severely limited, lengthy annotations lead to clutter, and the number of annotations that can be displayed on a video is limited, a problem that was identified by Hosack [32] as well. This approach is suited for platforms where only a single person is able to create annotations. In a social annotation environment, this approach is unfeasible: Here, only a set of annotations should be displayed on the video, e.g., the currently selected annotation or a set of filtered annotations. Text itself should also not be displayed directly on the video. The spatial component of the context should be the only element that is displayed directly on the video.

Another distinction - closely linked to context - lies within the display of annotations: When using a temporal context, the only viable option is to display annotations in close proximity to the video. This is used by Medithread and VideoANT. Those platforms display the annotations in a list next to the video. A click on an annotation starts playback of the associated point in time or time interval. Additionally, VideoANT highlights the current annotation when playing the video back.

Another interesting take-away from the previous section is the usage of controlled vocabulary: While a strict, controlled vocabulary is not desirable for a collaborative learning platform, the possibility to introduce a controlled vocabulary for parts of an annotation seems reasonable. An example would be peer grading where an annotation could offer in addition to a textual comment also the possibility to mark something right, i.e., a boolean value, and select a number of credits from a pre-defined range. This concept is not exclusive to video annotation and can be used with every kind of medium.

4.1.3 Image Annotation

Annotating images is another important aspect of a collaborative learning platform: Much of the content relevant for such a platform comes in form of images, such as digitalized artwork, maps, graphs, or diagrams. The rise of smartphones and the accompanied availability of cameras opens up new possibilities of interaction in a lecture: Students could take pictures of the whiteboard and immediately create annotations on the taken picture.

This section introduces various tools and platforms that allow the annotation of images.
4.1.3.1 EverNote

EverNote\textsuperscript{19} is an annotation platform that allows users to create so-called notes. Notes are basically documents with title and content. Notes can be organized in so-called notebooks. Notes and notebooks can be shared with other users which allows them to view or edit notes. The content of notes is not limited to text - users can add attachments, such as PDF documents or images, and use a rich text editor to create textual content.

EverNote allows the creation of annotations for attached PDF documents and images. EverNote treats every medium the same way: When annotating text documents, it is not possible to interact with the textual content of the documents, i.e., it is not possible to select passages of text for annotation. This behaviour resembles image annotation, and therefore, EverNote is classified as an image annotation platform.

\textbf{Figure 22: Annotations in EverNote:} Highlights using the transparent freeform drawing tool (top left), a textual remark (top right), shapes (middle left), a marker pointing on a passage of text containing a textual comment (middle right), arrows pointing at parts of the document (bottom left), and a freeform drawing using the opaque setting (bottom right). Annotated medium: [44].

The different kinds of annotations available in EverNote can be seen in Figure 22. One type of annotation is a freeform annotation which allows to draw arbitrary things on the medium. There are two settings for freeform drawings: Opaque and transparent. The opaque setting can be used to draw arbitrary things on the medium, shown

\textsuperscript{19} https://evernote.com
in the bottom right image of Figure 22. Highlights can be realized using the transparent setting, as can be seen in the top left image of Figure 22.

The top right image of Figure 22 shows a text annotation which is a piece of text added at an arbitrary position of the medium. Shape and arrow annotations, as can be seen in the middle right and bottom right image of Figure 22 respectively, can be used to add shapes, such as rectangles, circles, or arrows pointing at something, to the medium.

Figure 23: Marker icons available in EverNote.

A marker annotation adds an icon to the document. The available markers can be seen in Figure 23. Optionally, a piece of text and a triangle pointing in one of eight directions can be added to the marker. This can be seen in the middle right picture of Figure 22.

As already mentioned, the biggest drawback of annotating with EverNote is the missing interaction with the textual content of documents: Annotations exist on another layer above the text and can not refer directly to a passage of text. Another problem is the text annotation tool because textual comments are added directly onto the medium. While this approach mimics the behaviour of physical annotation more closely, it severely limits the length of textual annotations and introduces clutter to documents. The difference to physical annotation lies within the formatting possibilities: With physical annotation, users can format the textual content of annotations in a way that takes the content of a document into consideration, while with digital annotation the user is restricted to a passage of text in a rectangular text box.

Created annotations have spatial context because they can be placed anywhere on the document. EverNote allows textual and graphical annotations - textual annotations can be placed directly onto the medium via the text annotation tool, graphical annotations either through shapes or the freeform drawing tool. Annotations are informal and can either be shared or shared with other users making them public.

4.1.3.2 Marqueed

Marqueed\(^{20}\) is a web application that allows its users the collaborative annotation of images. To start annotating, users have to add collections and add images to collections. Collections can either be private or public. Images in a public collection can be viewed and annotated by every user, while images in a private collection can only

\(^{20}\) https://www.marqueed.com
be annotated by the owner and other users explicitly added to the collection.

Figure 24: Annotations on marquee: A rectangular selection with a textual comment (left image) and a closed polyline with a textual comment (right image). Annotated medium by Susan Blake.²¹

Users can define rectangular areas and arbitrary-shaped polylines on images and attach textual comments to them: Rectangular regions require a comment, for polylines comments are optional. Each polyline is seen as a separate annotation, i.e., it is not possible for an annotation to consist of more than one line. Figure 24 shows the two types of regions available on Marquee: The left image shows a rectangular selection around the pistil of the flower, and the right image a polyline with a textual comment. As soon as a textual comment exists for an annotation, other users can comment on the annotation.

Marquee offers spatial context for annotations. Annotations are either private or public, informal, and textual.

4.1.3.3 LabelMe for Images

LabelMe²² is an image annotation tool developed by the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. Goal of the tool is to create a corpus of annotated pictures to be used to train and evaluate computer vision systems. For the annotations to be used for computer vision systems, the annotations have to be formal. [71]

As shown in the left image of Figure 25, users annotate polygonal-shaped regions of images. In this case, a region for every petal of the flower was defined. The right image of Figure 25 shows the dialogue window for creating the description of a region: Users have to enter an object name, if the object is occluded, i.e., if every part of the object is visible or some parts are occluded by other objects, and attributes,

²¹ Available online: https://www.flickr.com/photos/forest_service_southern_region/18929553608
²² http://labelme.csail.mit.edu/Release3.0
such as the colour of a petal. Torralba et al. [71] mentions a property closed for an object door as another example.

LabelMe allows the definition of part-whole-relationships between objects of an image to further describe the structure of images: In the example in Figure 25, it would be conceivable to create a region blossom and add regions containing petals and the pistil as parts of the region blossom.

With LabelMe, annotations are inherently formal which is in line with the goal to create a training and evaluation set for machine learning algorithms. The annotation type is textual. Annotations are public, and context is spatial.

4.1.3.4 Bounce

Bounce\(^{24}\) is a web application that offers its users the possibility to create and share annotations for websites. To begin annotation, users enter an URL whereupon Bounce creates an image of the website which can be annotated. Annotations can be shared via link - users that have the link can view annotations and create own annotations.

Annotation means are similar to the other platforms examined in this section with the exception that only rectangular regions can be annotated. Figure 26 shows a part of a screenshot of a website with two annotations by two different persons indicated by different border colours. Clicking on a region displays the textual comment associated with the region.

Bounce offers spatial context and support for textual annotations. Created annotations are informal and either private or public.

\(^{23}\) Available online: https://www.flickr.com/photos/forest_service_southern_region/18929553608

\(^{24}\) http://www.bounceapp.com
4.1.3.5 Conclusions for Image Annotation

The previous section gave an overview of different image annotation platforms, the kind of annotations they support, and the different ways they utilize to display the annotations.

One thing supported by every platform was spatial context, i.e., all of the platforms gave users a way to create regions on the image the respective annotation is referring to. The main difference lies in the form of the region.

Rectangular regions are both easy to use and easy to implement, but there are cases in which a rectangular region is not accurate enough, e.g., when annotating the borders of countries on a map which are seldom rectangular. If the context of a collaborative learning platform needs polygonal-shaped regions remains to be seen, but polygonal-shaped regions should be considered when implementing image annotation. Some platforms do not require a polygonal region to be closed which was called a polyline. A polyline can be seen as a graphical sign, but an implementation - as opposed to the implementations found on most examined platforms - should allow a graphical sign to consist of more than a single polyline. The pros and cons of graphical signs are the same as those of graphical signs used for annotating textual media and can be found in Section 4.1.1.7.

Another type of selection was a point selection, i.e., annotating a single point on an image which can be found on Mediathread.

With the exception of Mediathread, all of the examined platforms where collaboration was possible displayed every annotation on the image which leads with already a small number of annotations to clutter, even if only the region and not the annotation itself is displayed. Looking at one of Marqueed’s more annotated images, the cluttering of the image is clearly visible while only showing the regions of annotations. A collaborative learning platform should offer filter options for annotations so that users can control which annotations are displayed for them.

---

25 http://www.pms.ifi.lmu.de
26 https://www.marqueed.com/collections/2956/images/11679
Most of the examined platforms only supported textual annotations associated with a region, but EverNote allowed users to place freehand drawings on the medium. Here has to be considered that a graphical annotation adds clutter to the medium, but there are cases where freeform annotations are useful and more appropriate than a mere textual annotation: An example would be an image from a whiteboard where a minus is missing: Instead of creating a textual annotation with content “-”, one could simply draw the missing minus directly on the image.

EverNote offers marker annotations which are basically images optionally associated with short a textual comment. Markers can be seen as a textual comment with a purpose, e.g., checks can be used to signify that the passage the marker is pointing to is correct. On a collaborative learning platform, this could be used, e.g., in peer grading: A check with achieved score as textual comment to signify a correct task, a cross for an incorrect task.

4.1.4 Audio Annotation

Lecture recordings, podcasts, or music - audio content is another type of media that has to be considered for annotation. Audio content is very similarly to video content with the difference that there is no spatial component to it. Considering that most of the examined video annotation platforms merely supported temporal instead of spatio-temporal context, many findings for video annotation can be applied without major restrictions to audio annotation.

Admittedly, there are possibilities to create a spatial dimension for audio items by visualizing the content, e.g., in form of a wave diagram. This visualization could then be annotated to create annotations with spatio-temporal context. Use cases for visualization and subsequent annotation are generally of a more scientific nature, e.g., the mapping of morphemes to a wave diagram.

4.1.4.1 SoundCloud

SoundCloud\textsuperscript{27} is a social network for music. On SoundCloud, artists can share their self-produced music tracks for other users to rate and comment on. A music track on SoundCloud is visualized as a wave diagram, but comments are only possible for a point in time, so context of annotations is temporal. Comments on SoundCloud are always associated with a point in time, so it is not possible to create a comment for a whole music track.

Figure 27 shows the display of annotations on SoundCloud: The vertical bars form a wave diagram of the music track, the colours orange and white are used to represent playback progress. The pictures

\textsuperscript{27} https://soundcloud.com
in the bottom part of the image are the profile pictures of annotators and are displayed at the point in time the respective user created an annotation. When playing back a music track, an annotation for the current playback progress is displayed below the visualization if one exists.

While one could easily conclude that annotations on SoundCloud use temporal context, an observation of existing annotations shows that most annotations are definitely intended as a comment on the whole music track. This annotation behaviour can possibly be explained by the way comments are created on SoundCloud: The user interface does not make it clear that a comment is created for the current playback progress but rather offers a simple text area, which can also be found, e.g., on YouTube. But there are also exceptions where users comment on a specific point in time, e.g., “love this part!!!”.

Regardless of this observation, context offered by SoundCloud is temporal. Created annotations are textual, informal, and public.

4.1.4.2 Sonocent Audio Notetaker

Audio Notetaker\textsuperscript{28} is an example for commercial audio annotation software. It is mainly intended for recording lectures or meetings and creating annotations for those recordings.

Figure 28: Annotation interface of Audio Notetaker: Horizontal compartments are used to add additional information to a set of chunks, while vertical compartments are used to structure the chunks into set of chunks.

The recording is visualized as a list of horizontal bars, called chunks, where each chunk represents a part of the recording where there is sound. The annotation interface of Audio Notetaker can be seen in

\textsuperscript{28} https://www.sonocent.com/en/audio-notetaker
4.1 BENCHMARKING ANNOTATION PLATFORMS

Figure 28: There are horizontal (Images, References, Text, and Audio) and vertical (yellow and green background) compartments. The chunks are the bars in the rightmost horizontal compartment captioned Audio. At the beginning, the recording is a chronologically ordered set of chunks which then can be split into smaller sets of chunks. Splitting chunks leads to the creation of a new vertical compartment. In Figure 28, the recording was split into two sets. A conceivable use for splitting chunks into sets is to structure the recording, e.g., by slides or semantic units. It is also possible to change the order of chunks, i.e., the structure is not necessarily a chronological ordered one. For this to work effectively, a chunk should represent a whole sentence or passage. This requires the annotator to rectify the chunks by merging or splitting them because the software does not always correctly recognize the beginning and the end of sentences.

Chunks can be given a colour which allows users to code the recording on two different levels of granularity. Main reason for colouring is that Audio Notetaker allows users to create a new recording containing only chunks of a certain colour. This feature allows users who coloured all chunks for a certain topic in a single colour to create new recordings about a certain topic.

As can be seen in the right part of Figure 28, annotators can add images, references, and own notes to a set of chunks. A possible use case would be to split the lecture into sets of chunks where every set represents a single slide of the lecture. Afterwards, an image of the slide and own notes can be added to the respective set of chunks.

After annotating, the mix of notes, images, and audio can be displayed in a single window which potentially allows a more effective review of the recording.

Context is temporal, and the annotations are private, but the recording and the associated annotations can be exported as a file which then can be shared with others. Created annotations are informal. Audio Notetaker supports textual and graphical annotations by adding an image to a set of chunks. The horizontal compartment captioned Reference in Figure 28 is misleading: It is intended to separate references, such as links, from one’s own notes but not to create references to other parts of the recording or other media.

4.1.4.3 Soundmarks

Soundmarks\textsuperscript{29} is an application for iOS devices allowing the simultaneous recording and annotation of audio content.

After starting recording, annotators can add one of three different types of annotations to the current point in time: Text, freeform drawings, and images. Annotations are associated with the point in time at which the user started to create the annotation. Annotations are

\textsuperscript{29} https://itunes.apple.com/de/app/soundmarks/id625402658
displayed in a chronologically ordered list, as can be seen in Figure 29. From top to bottom, there is a text annotation, a freeform drawing annotation, and an image annotation. It is also possible to add annotations afterwards while playing the recording back - here, annotations are added to the current playback progress.

When playing the recording back, there is no indication which annotation is associated with the current playback progress. Users either keep track of the current annotation by comparing timestamp of the recording with the timestamp of the annotations or tap on an annotation which leads to playback starting from the timestamp associated with the annotation.

The created annotations are private, but the user has the possibility to export the recording plus annotations which can be shared with other users. The context is temporal and the annotations are informal. Annotators can create textual and graphical annotations.

4.1.4.4 Conclusions for Audio Annotation

In this section, various platforms supporting the annotation of audio content were introduced. As already mentioned, the annotation of audio content is very similar to the annotation of video content, so most of the conclusions for video annotation also apply here.

Besides for very short audio content, temporal context for annotation of audio content is inevitable. As with video content, there are two possibilities for temporal context: Annotating points in time or time intervals. While the former possibility is faster to create, the latter one offers a more precise definition of the part of the medium the annotation is referring to.
One of the biggest challenges - same as with other media - is displaying the annotation in a social annotation environment. There is no problem when annotating privately because there are only annotations of a single person to display. In a social annotation environment, there is more than a single annotator and possibly more than one annotation associated with a point in time or interval which raises the question which annotation to display and how to handle overlapping intervals?

4.1.5 Annotations in Social Networks

The emergence of social networks in the last decade had serious impact on the way users interact with websites - the interaction got more social and more focused on content created by users themselves. Many of those interactions can be seen as implicit annotations created by users - implicit, because the action the user did had another goal than creating an annotation.

Generally, there are five concepts observed in social networks which can be used to create implicit annotations. It is certainly possible and likely that there are concepts in social networks that are not mentioned in this section, but the concepts mentioned here are the most prevalent. The annotations created by those interactions are mostly relevant for the latter part of this thesis where different forms of displaying and filtering of annotations are explored. This section introduces the concepts likes, up- and downvotes, ratings, sharing, and mentions.

4.1.5.1 Concept “likes”

The concept likes can be found on many social network sites, such as Facebook\textsuperscript{30} (in form of likes), Twitter\textsuperscript{31} (in form of favoriting a tweet), Instagram\textsuperscript{32} (in form of hearting a picture), or Google Plus\textsuperscript{33} (in form of +1’ing). Contrary to popular belief, the like feature was not invented and spearheaded by Facebook but by FriendFeed in October of 2007. Facebook’s like button was not launched until almost two years later in February 2009.\textsuperscript{[7]}

Liking something is the expression of appreciation for something, e.g., photos on Instagram or posts on Facebook. For a collaborative learning platform, and more generally speaking for any social platform, not implementing the concept likes is inconceivable. The real question is on which level of granularity the concept should be implemented: On annotation level, on medium level, or on lecture level?

---

\textsuperscript{30} https://www.facebook.com
\textsuperscript{31} https://twitter.com
\textsuperscript{32} https://www.instagram.com
\textsuperscript{33} https://plus.google.com
One could argue that likes can be seen as binary ratings, but such an interpretation has to be treated with some reservation. While the presence of a like definitely expresses appreciation, the absence of a like can not be equated with dislike. As soon as it is possible to dislike something, the interaction is covered by another concept - up- and downvotes.

4.1.5.2 Concept “up- and downvotes”

As already mentioned, the concept up- and downvotes is very similar to the concept likes with the difference that it is possible to dislike content in form of a downvote which is why this concept may be seen as binary rating. Up- and downvotes can be found on social networks such as reddit\(^{34}\), Delicious\(^{35}\), or Digg\(^{36}\).

Figure 30: Up- and downvotes on reddit: The score in the middle is the current score of the post, the arrows above and below can be used to give an up- or downvote respectively.

Figure 30 shows up- and downvotes on reddit: The score in the middle of the two arrows represents the current score of the post, and the two arrows can be used to up- or downvote the post. In case of reddit, the score is the difference between up- and downvotes or zero if the difference is negative.\(^{37}\) With the default setting, posts on reddit are ordered by another score, the hot score: The hot score uses the logarithmic value of the score and adds a factor that is higher the more recent the post is.\(^{38}\)

Independent from the social network, content with a high score should represent well-liked content, while content with a low score should indicate disliked content. A similar but more general concept are ratings.

4.1.5.3 Concept “ratings”

While ratings as a concept is older than social networks it is found on many social network sites and is, therefore, considered here. A rating is a value from a fixed scale or range associated with a content expressing the subjective degree of appreciation for the respective content. An example for a scale is a five star scale, with one star

\(^{34}\) https://www.reddit.com

\(^{35}\) https://delicious.com

\(^{36}\) http://digg.com

\(^{37}\) https://github.com/reddit/reddit/blob/master/r2/r2/lib/db/_sorts.pyx, lines 41-42

\(^{38}\) https://github.com/reddit/reddit/blob/master/r2/r2/lib/db/_sorts.pyx, lines 47-58
representing extreme dislike, five stars representing extreme like, and the stars in between representing different degrees of like or dislike respectively. In a social environment, the ratings of different users are accumulated in a single score which can be calculated in different ways, the easiest one being the mean value of all ratings.

Companies and businesses can represent themselves on Facebook using a so-called business page. If a business can be visited by people, e.g., a restaurant, bar, or a university as opposed to a car manufacturer, the check-in feature can be enabled which allows users to create Facebook postings with a reference to the business page. An example would be the Facebook page of LMU Munich. As soon as the check-in feature is enabled, the business page, and in extension, the business itself can be rated on a five-star scale with an optional textual comment.

Other social networks, such as IMDb, Rotten Tomatoes, or Goodreads allow their users to rate the kind of content the website is covering, in this case, movies and books, respectively.

4.1.5.4 Concept “sharing”

Sharing, which allows users to share content with other users of the network, is a concept that has gained foothold in many social networks. Depending on the network, shared content is from the network itself, such as a posting of a user, external content, such as an Amazon product page, or both.

Facebook supports both kinds of sharing: Users can share Facebook posts which leads to a copy of the post, optionally with a comment, appearing on the sharer’s own timeline. External websites can include a share button which allows users to share the link to the website with an optional comment.

Retweeting on Twitter can also be seen as a form of sharing: Retweeting a tweet adds the tweet and an optional comment to the retweeter’s timeline. Twitter offers external websites the option to include a Twitter button which has the same functionality as Facebook’s share button: It creates a tweet with a link to the website optionally with a comment.

The reblogging feature on Tumblr is an example for a social network where sharing is restricted to content that already exists on the social network. Tumblr is a so-called microblogging platform that allows its users to create blog posts containing different kinds of media, such as videos, images, or text. Users can reblog content found on other blogs on Tumblr which leads to a copy of the blog post appear-

39 https://www.facebook.com/lmu.muenchen
40 http://www.imdb.com
41 http://www.rottentomatoes.com
42 https://www.goodreads.com
43 https://www.tumblr.com
ing as post in their own blog. Optionally, a comment can be added to the reblogged post. Reblogging is one of the main concepts of Tumblr: It can be seen as an easy option to add new content to one’s own blog.

4.1.5.5 Concept “mention”

The concept mention describes a hypertextual reference to an entity on a social network, such as a person or a company represented by an account on a social network. The established notation for a mention is an @-sign followed by the account name. Most social networks offer autocompletion after the user entered a few characters allowing to select the target of a mention from a list which ensures that the mentioned entity exists on the social network.

Figure 31 shows mentions and the associated autocompletion on Facebook and Twitter. In both cases, the suggestions for LMU are shown. Many social networks, such as LinkedIn, Tumblr, or XING, offer their users the possibility to create mentions.

Figure 31: Mentions on Facebook and Twitter with autocompletion for the term LMU.

An interesting point of mentions is that there are two different semantics: The original semantic, most probably emerged in the early days of the bulletin boards, is addressing a person by using the @-sign followed by the name of the person to make clear the the following content is meant for this person. The other semantic, emerged in recent years, is using a mention for creating a hypertextual reference to the person, e.g., the Twitter account of a news station using @barackobama in a tweet promoting an article about Barack Obama. The difference is that in the first case the content is addressed to the mentioned entity and in the second case about the mentioned entity. The two kinds of mentions express different relationships: The first kind expresses a relation between two entities, while the latter expresses a relation between a piece of content and an entity.

44 https://www.linkedin.com
45 https://www.xing.com
4.1.5.6 Conclusions from Annotations in Social Networks

The different concepts of interaction found on social networks can be used to create implicit annotations. Implicit, because the goal of users is not to create an annotation but, e.g., liking or sharing content on a social network. As already mentioned, implicit annotations can be used for background tasks, such as determining ranking measures for annotations. This is important for a collaborative learning platform because there is a great number of annotations that have to be ordered and displayed.

Likes, up- and downvotes, and ratings are semantically closely related: All are used to give users the possibility to express their subjective liking of content. Because of their similarity only one of the three concepts should be implemented to create a standardized and consistent way of expressing like, as proposed by user interface design principles. The U.S. Department of Health & Human Services proposes in their usability guidelines to “allow users to perform tasks in the same sequence and manner across similar conditions.” [64] Shneidermann and Plaisant propose to “strive for consistency” [66, p. 74] as one of the eight golden rules of interface design. Therefore, there are two questions: Which of the three concepts to implement and on which level of granularity - annotation, medium, or lecture?

Mention is a concept that is not necessarily required for a collaborative learning platform but a nice addition. If annotations can be commented on, users can use mentions to address different questions of different people in a single comment. This would conform to the address semantic of a mention. The two different semantics have implementation-wise no difference, and therefore, users are free to choose whichever they want to use.

For the current vision of the collaborative learning platform, sharing content is not required because there is no timeline or other medium where the shared content could be displayed.

Another concept often found on social network sites is tagging, but because of the more complex nature of tagging the whole next section is devoted to this concept.

4.1.6 Tagging

The Oxford Advanced Learner’s Dictionary defines a tag as a “label attached to someone or something for the purpose of identification or to give other information.” [31] Tagging, therefore, is the process of attaching a label to something which are, in case of a collaborative learning platform, different kinds of media and annotations themselves.

Mathes [51] identifies three different kinds of tagging: Expert tagging, author tagging, and social tagging. To be exact, Mathes is not talking about tags but about metadata in general. But, as tags are
a form of metadata this classification can be applied to them. Two kinds of tagging that emerged later and are therefore not mentioned by Mathes are crowd tagging and hashtags. For this thesis, crowd tagging is seen as tagging using the wisdom of the crowd. Hashtags are seen as tags that are included into the actual comment text by being prefixed with a #-sign.

The following section introduces the five kinds of tagging in detail and illustrates them with examples.

4.1.6.1 Expert Tagging

One can expect that tags created by experts are of high quality but are both expensive and time-consuming to create. An example for expert tagging can be found in most libraries where librarians create tags for library media. Metadata created by experts is often done by following rules for structure and a controlled vocabulary. An example proposed by Mathes for a controlled vocabulary is the Library of Congress Subject Headings, a vocabulary to be used “to catalogue the materials held at the Library of Congress.” Mathes cites the MARC21 standard as an example for a set of structure rules for creating metadata. The created metadata is machine-readable and exchangeable between other libraries or other parties which use the MARC21 standard. In the context of a collaborative learning platform, the person creating content are often lecturers themselves - which can be regarded as experts in the area they are creating lectures and content in - so that tags created by the lecturers can be seen as expert tags.

4.1.6.2 Author tagging

The idea of author tagging is that creators of content create tags that are used to classify and identify content. This approach can be found on platforms such as YouTube or Tumblr where users adding content to the platform have the possibility to add tags to their content. These tags can be used to make the content searchable.

A disadvantage with this approach, as stated by Mathes, is that “the intended and unintended eventual users of the information are disconnected from the process [of creating tags].” In other words, the author of the content uses his tags to describe the content - tags that are not necessarily the tags an eventual consumer would use to describe the content and, therefore, not appropriate to search for the content. A possible explanation for this behaviour, suggested by Golder and Huberman, are different levels of experience or domain knowledge between author and consumer. For example, technical terms such as instance variable or database query are known to experienced but not to inexperienced users. A conceivable relief for
this problem could be the usage of a controlled vocabulary for tags which trades a loss in expressivity for a gain in conformity.

To further close the gap between author and consumers, the whole tagging process can be made social.

4.1.6.3 Social Tagging

The difference between social and author tagging is that with author tagging there is a single authority in charge of the tags, while with social tagging this authority consists of all people annotating a certain content. [51] Van der Wals calls “the result of personal free tagging of information and objects (...) for one’s own retrieval” [73] a folksonomy - a neologism of folk and taxonomy.

Users of social tagging platforms typically create their own tags for a resource. Tags created by the users of the platform for a single resource are lumped together and form the set of tags describing the resource. Depending on the platform, this set of tags is either visible to users or just used for retrieval and ordering of results. [26]

An example for a social tagging platform that only displays tags of a single user is Delicious which calls itself a social bookmarking service. As implied by bookmarking, the tagged resources on Delicious are bookmarks. The motivation of users for tagging bookmarks is the organisation and classification of their bookmarks. Users can browse the bookmarks of other users or search for bookmarks tagged with a certain tag.

Foursquare is a social network that lets users explore interesting places in their surroundings, such as restaurants, bars, or shops. There are two sources for tags on Foursquare: Users can add a tip - a textual comment - to a location. Afterwards, tags are extracted from those tips. An example for tag extraction can be seen in Figure 32: The terms surrounded by bubbles are tags, and when clicking on a tag, the associated tips are displayed below. The other source of tags can be seen in the right image of Figure 32: When rating a place, Foursquare not only offers a numerical scale but also asks questions about the place - answering the question in the image with very would lead to the creation of another instance of the tag authentic.

Goodreads, a social network about books, allows users to organize books in so-called shelves. Adding a book to a shelf can be expressed in tag terminology as adding a tag to a book. The shelves of a book are used by Goodreads to determine the genres of the book: The ten shelves most users filed the book in are used as the genres of the book. Not all shelves are considered as genres, such as the shelves to-read, need-to-buy, and kindle-ebooks. How Goodreads identifies shelves that are used for genres is not apparent. One conceivable

---

46 https://delicious.com
47 https://foursquare.com
48 http://www.goodreads.com
possibility would be that the shelves for genres are predefined which would be tantamount to using a controlled vocabulary.

4.1.6.4 Crowd Tagging

Another approach for tagging was introduced by Ahn and Dabbish [72] with the ESP game: In the ESP game, tags are not only collected but also validated by disguising the tagging process as a game. Here, two players who are shown the same image have to enter possible tags for the image: As soon as both players have a match, i.e., they entered the same tag, this tag is seen as a valid tag for the image and the players are shown the next image. A game round consists of fifteen pictures and has a duration of 150 seconds. [72]

A modification of the ESP game called ARTigo game can be found on the ARTigo platform49 - here, two players are shown an artwork for a fixed duration of sixty seconds and are entering tags the whole time. A matched tag does not lead to the next picture but earns the players points. The next image is shown after sixty seconds, and a game round consists of five pictures. Additionally, the ARTigo platform offers other games that aim for others things in the context of tagging, such as refining tags or identifying important tags. [76]

4.1.6.5 Hashtags

A hashtag is a word prefixed with a hash sign (#) and can by now be found on many social network sites, such as Twitter, Facebook, or Instagram. Chris Mesina is considered the inventor of the hashtag, first

---

49 http://www.artigo.org
proposed in a tweet (Figure 33) and his blog. The use of the hash sign for groups is borrowed from IRC where channels on a server are prefixed with a hash sign. [57] While Mesinas proposal dates back to August 2007, Twitter first started to officially support hashtags in July 2009 by converting hashtags to links that lead to an overview page showing other tweets using that hashtag.

Figure 33: First use of a hashtag (color of the hashtag changed to resemble the original tweet of 2007 more closely).

Bruns and Burgess [9] identify three usage patterns of hashtags on Twitter. The first usage pattern is “to mark tweets [that] are relevant to specific known themes and topics” [9, p. 3], such as #worldcup2014 for the 2014 FIFA World Cup or #mh370 for the disappearance of the Malaysia Airlines Flight 370. Bruns and Burgess call hashtags in such a context thematic hashtags. Their interpretation is that a thematic hashtag is used “to address an imagined community of users who are following and discussing a specific topic.” [9, p. 4]

Thematic hashtags are relatively specific and another observable usage pattern is the usage of generic hashtags, such as #Deutschland or #Twitter. Bruns and Burgess offer two possible interpretations: The first is tweeting about breaking events for which so far no specific hashtag exists or as “a simple means of emphasis.” [9, p. 5] Emphasis is required because Twitter does not allow any kind of formatting in tweets, but hashtags are coloured differently and, therefore, may be used as an alternative for formatting. [9]

The last usage pattern are called empathic uses by Bruns and Burgess used to “express the sender’s emotional or other responses” [9, p. 5], such as #swag, #happy, or #yolo. A usage pattern related to empathic uses are Twitter memes, which are empathic uses that are followed by people, i.e., are again addressing a group, such as #firstworldproblems or #fail. [9]

Another usage pattern that could be observed in the current European refugee crisis is using a hashtag to send out a signal - in this case #refugeeswelcome. This uses Twitter’s trending hashtags where the most popular hashtags are displayed on every Twitter page, i.e., if enough people create tweets containing #refugeeswelcome, it is displayed on every Twitter page.

50 http://factoryjoe.com/blog/2007/08/25/groups-for-twitter-or-a-proposal-for-twitter-tag-channels
4.1.6.6 Conclusions from Tagging

Tagging is a very important type of annotation: Tags can be used to make content on a collaborative learning platform searchable without only resorting to information retrieval techniques using the textual content available on the platform. An incentive for users to tag content is organisation: Tagging content on the platform allows users to categorize and structure information.

As seen in the previous section, there are different types of tags that are of varying significance for a collaborative learning platform.

On a collaborative learning platform, expert tags are most likely tags created by authors of lectures which can be considered experts in their subject area. Lecturers should be required to tag the content they make available on the platform so that content can be made search- and retrievable.

There are different kinds of authors on a collaborative learning platform: Lecturers, which were already discussed, annotators, and commenter. Should an author be able to tag his content regardless of its type? Not mandatory, but the possibility should be there to support searchability of content.

Users should also be able to create private tags for content of other users, in the nature of social tagging, to give users the possibility to organize and structure content for further use. For this to work effectively, the platform has to offer capabilities to display, search, and filter content tagged with private tags.

The only usage pattern of hashtags that could potentially find a place on a collaborative learning platform are hashtags for addressing a group of people. Even then, the first usage pattern is not really required if the platform supports groups and different levels of privacy, such as setting an annotation only visible for a certain group. The other three usage patterns are of a more personal nature and are, therefore, not required on a collaborative learning platform.

4.1.7 Wrapping up Benchmarking Annotation Platforms

This chapter introduced annotation platforms for different types of media. Ideas and approaches to annotation for the examined types of media were taken into consideration for usage on a collaborative learning platform. Goal of this chapter was to work out different types of annotations for different types of media and which of them to implement on a collaborative learning platform.

Annotations on social networks were discussed: Social annotations are annotations created by interaction of users with a social network. These annotations are mostly implicit and can be used to control the display and ordering of annotations. They are important for any social annotation environment where the number of annotations could
potentially overwhelm users. Filtering and ordering are discussed further in Chapter 5.

Tags are important for search and retrieval but a mighty tool for users to organize and structure information found on the platform as well. Finding a lecture about logic without any of the lectures tagged with appropriate tags may prove difficult but as soon as tags are introduced this task becomes easy. Organizing and structuring are important when users work with content on a platform: Without a possibility to mark content relevant for a certain topic, the user would be forced to start memorizing where the information is found or, more possibly, stop using the platform at all.

Chapter 3 introduced different types of annotations and the platforms were classified by these criteria. The exact definition of the classification criteria and possible extensions for the classification framework are discussed as the last part of this chapter.

4.2 CREATING A FRAMEWORK FOR ANNOTATIONS

This part of the chapter uses the results from the benchmarking and the theoretical foundation laid in Chapter 3 to develop a classification framework for annotation platforms. Not every distinction proposed in Chapter 3 is part of the framework: Reason for that is that there are classifications one can not make without knowledge of the background workings of the platform, e.g., if a platform supports or uses implicit annotations.

Context of an annotation is an important part of the classification framework because benchmarking has shown that even for a single type of medium platforms support different types of context.

(I.) An annotation platform is classified as supporting spatial context if an annotation can be placed at a spatial position in medium.

(II.) An annotation platform is classified as supporting temporal context if an annotation can be associated with a temporal segments of a medium.

(III.) An annotation platform is classified as supporting spatio-temporal context if (I) and (II) hold for the annotation platform.

The distinction between referencing, textual, and graphical signs is another classification that is important for the framework because representation is an important aspect of annotations which would get lost without such a distinction. Benchmarking showed that this classification is not fine enough: A great number of platforms allowed annotators to highlight passages of text or parts of a medium without giving them the possibility to create freeform drawings. To high-
light this difference, graphical annotations are split into freeform and highlight annotations.

(I.) An annotation platform is classified as supporting textual annotations if the content of an annotation can be of a textual nature.

(II.) An annotation platform is classified as supporting referencing annotations if the content of an annotation can be a hypertextual reference to another medium or part of the same medium. Links in a textual annotation are not counted as referencing annotations.

(III.) An annotation platform is classified as supporting freeform annotations if the content of an annotation can be free in form, possibly only limited by choice of colour or other attributes. Freeform annotations can be comprised of more than one line.

(IV.) An annotation platform is classified as supporting highlight annotations if an annotation can highlight parts of a medium.

Public and private is another distinction that plays an important role for the framework because it is a first indicator for a platform being socially-oriented or mainly intended for private annotation.

(I.) An annotation platform is classified as public if annotations can be made available to any user of the platform.

(II.) An annotation platform is classified as private if annotations are created for private use and cannot be shared with other users of the platform.

Another distinction is between formal and informal annotations. Formal annotations can be easier facilitated for search and retrieval than informal annotations.

(I.) An annotation platform is classified as supporting formal annotations if an annotation can follow a predefined structure.

(II.) An annotation platform is classified as informal if it is not a formal annotation platform.

Omitted were the distinctions between apparent and non-apparent, explicit and implicit, and permanent and transient annotations. As already mentioned, the distinction between explicit and implicit annotations is not part of the classification framework because one cannot observe from the outside if a platform creates implicit annotations or not. Permanent and transient are omitted because there is absolutely no reason for a digital platform to make annotations permanent and the inclusion would only serve as a “call-out” to platforms that make
annotations permanent. Apparent and non-apparent did not find a place in the framework because this classification is not a static, but a dynamic one: It depends on how users use the tools given to them and, therefore, is not considered here.

A possible extension for the framework is media richness of annotations, i.e., what kind of media can be used in an annotation, e.g., creating an annotation for a passage of text in form of a video. Alas, none of the examined platforms offered capability to this end, and therefore, this classification is not considered here but outlines a plausible extension for the classification framework.

The classification framework introduced in the previous section is now used in Table 2 to classify the examined annotation platforms.
## Classifying the Annotation Platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>Context</th>
<th>Sign</th>
<th>Privacy</th>
<th>Formality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>spatial</td>
<td>temporal</td>
<td>spatio-temporal</td>
<td>textual</td>
</tr>
<tr>
<td>Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YouTube</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>VideoANT</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Mediathread</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>ANVIL</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrobat Reader</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>iBooks</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Google Docs</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Kindle</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Annotate.co</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>AnnotateIt</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LabelMe</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>BounceApp</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>EverNote</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Mediathread</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>marqueed</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Audio</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SoundCloud</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>AudioNotetaker</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Mediathread</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Soundmarks</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

Table 2: Classification of the examined frameworks.
4.2.2 Wrapping up Creating a Framework for Annotations

In this section, the theoretical foundation from Chapter 3 and the results from the benchmarking were brought together to build a classification framework that can be used to classify annotation platforms. Not all distinctions from Chapter 3 were appropriate for the classification framework, and not all distinctions were extensive enough to cover the diversity in annotations found on the annotation platforms.

The framework is a starting point: It is by no means exhaustive and can be extended if the need arises. For this thesis, the framework was mainly required to find the point when enough platforms were examined. Nevertheless, the framework and the classification offer a solid overview of annotation means for different types of media and can be used for other purposes.

Enough platforms were examined because the coverage of the different classes per media type in Table 2 is sufficient:

- For each type of context (if more than one is applicable), at least two platforms were examined.
- For each sign, except for referencing annotations, at least one platform was examined.
- For each characteristic of privacy, at least two platforms per type of media type were examined.

Two exemptions for formal and referencing annotations were made. For formal annotations, the reason for is simple: Formal annotations are important for platforms where the focus lays on creating annotations for retrieval or for scientific topics, such as linguistic analysis of videos. On collaborative learning platform, it is likely that most of the annotations are informal with formal annotations being the exception.

For referencing annotations, a strict definition was chosen: For an annotation to be considered referencing, its content had to be comprised of a hypertextual reference to other parts of a medium. This led to very few platforms supporting the referencing annotation which meets this definition which is why the criteria of having at least one of each sign per type of media was relaxed for referencing annotations.

The strict definition is required because otherwise a hyperlink inside a textual comment would count as a referencing annotation which is not the particular meaning of a referencing annotation. A referencing annotation is more than just a link - it is a link to a context. While links support a basic form of context in form of anchors\(^52\), those have to be defined on the respective website and are often not fine-grained enough or defined at all.

\(^52\) # in the URL followed by the name of the anchor
4.3 Wrapping up annotations “in the wild”

This chapter is the foundation for the remainder of the thesis: A great variety of annotation platforms and applications were examined in the first part of this chapter. After that, interaction concepts in social networks and tagging independent of a certain platform were examined. The interaction concepts in social networks can be used to create implicit annotations while tags are an important resource for search and retrieval.

Lastly, the results from Chapter 3 and the benchmarking were put together: The theoretical foundation was used to sketch the framework, the findings from the second part to extend the theoretical foundation. At last, the examined platforms were classified with the newly-created framework to decide if enough platforms were examined to regard the benchmarking as finished.

The next chapter tackles problems that arise from the social and collaborative orientation of the platform: A great number of annotations means that there have to be possibilities for filtering and ordering annotations so that users can still work effectively on the platform.
The digitalization of annotations opens up new opportunities for collaboration on a learning platform but also introduces challenges that have to be overcome before collaboration becomes completely viable. As already mentioned in the previous chapters, it can be assumed that with collaborative annotation there will be a greater number of annotations compared with private annotation, as well as annotations of greatly differing quality.

The following argumentation is similar to the argumentation Jan et al. [36] use to justify the development of an annotation filter for a collaborative annotation platform. Humans only have a limited amount of memory that can be used to process the current task at hand - this memory is called working memory. [14] The working memory “is used to plan and carry out behavior” [14, p. 325] In the context of working with annotations, it is used to process an annotation, its context, and a discussion revolving around the annotation. A related term is cognitive load, coined by Sweller [69], and is closely linked with working memory. Cognitive load describes how much of the working memory is used by the task at hand. [69] According to Dillenbourg [21], cognitive load is important for learning because “there is no learning without cognitive load, but there is no learning with too much cognitive load either.” [21, p. 142] Too much cognitive load is called cognitive overload. [21] Putting all this information together, it is easy to conclude that the cognitive load is higher the more annotations are visible to the user, and that it is important to devise mechanisms to keep the cognitive load at appropriate levels. This is backed by a study conducted by Wallen et al. [74] which examined the effects of having access to different types of annotations which differed in their elaboration: One kind only explained terms, another one explained whole ideas, and the last one connected the different ideas with each other. One of their findings was that having access to more than one type of annotation had a detrimental effect on learning performance. As a possible reason Wallen et al. cite cognitive overload. Pohl [62] also shares the opinion that the information flood has to be reduced, because “Computer-Mediated Communication (CMC) can quickly become confusing and incomprehensible, even when the number of participants is small,” [62, p. 89] and annotation in a collaborative setting can most definitely be seen as Computer-Mediated Communication.

The shown disadvantages could lead someone to believe that it is better for a learning platform to only offer means for private annotation. Chapter 2 already discussed upsides of having access to anno-
tations of others, and there is additional evidence that having access
to annotations that show different perspectives of the same problem
has a positive effect on learning. [30, 22]

Hooper and Rieber [30] state that more than a single perspective
is required to reflect on the complexity of most concepts and that
more perspectives are necessary “to mediate deep processing.” [30, p.
162] Dolmans et al. [22] agree with Hooper and Rieber that viewing a
problem “from multiple perspectives increases transfer of or the flex-
ibility with which learners can deal with new sets of events.” [22, p.
733] They argue as well that collaborative elaboration, which includes
note-taking, can have a positive effect on learning. [22]

For a collaborative learning platform, these results indicate that it
is important to give users access to annotations of other users, but
that the number of annotations should be limited to prevent cogni-
tive overload. Before someone can reduce the number of annotations,
different ways of displaying annotations have to be devised. For this,
Marshall et al. [49] again provide a solid starting point. After that, dif-
ferent ways to contain the information flood are discussed - either by
ranking annotations, which gives more important annotations prece-
dence before less important annotations, or by filtering, which hides
less important annotations completely for the user. First research in
this area was done by Jan et al. [36] who developed two kinds of
annotation filters using implicit feedback.

5.1 DISPLAYING ANNOTATIONS

Most platforms examined in Chapter 4 displayed all available anno-
tations directly on the medium. There were exceptions, such as An-
notate.co, which offers filtering possibilities, or Medithread, which
displays only a single annotation at a time.

Marshall et al. [49] identify three different ways to display annota-
tions on a digital reading device: Directly on the medium, as aggre-
gates of annotations, or detached from the medium. Their findings
refer to textual media but are adapted to image and video in this
section.

Figure 34 shows the three different display possibilities for display-
ing annotations on a textual medium proposed by Marshall et al. [49]
The left image shows the approach adopted by most of the exam-
ined annotation platforms: Display all annotations of all authors on
the document. The image in the middle shows the second approach
which merges annotations together to show parts of the document
where annotations accumulate. In the example, two different shades
of grey are used to identify passages with different annotation den-
sity. A similar approach for colouring aggregated regions has been
pursued by Shipman et al. [65] A real-life example of this approach
can be found on Amazon’s Kindle which highlighted passages that
were annotated by a certain amount of readers. The last approach can be seen in the right image: The annotated passages of text (in grey) and the respective comment (in black) are displayed detached from the text. The bottom two items in the list are annotations without comment, e.g., a highlight without comment. This approach was often employed besides displaying the annotations directly on the document, e.g., by Amazon’s Kindle and Apple’s iBook which offer an annotation overview structured in a similar way.

As already discussed in the introduction to this chapter, the first and third approach have to be used in conjunction with means for filtering and ranking annotations to prevent cognitive overload of users. The second approach already reduces cognitive load because by creating aggregates of annotations the number of annotations displayed on the document is reduced.

While Marshall et al. [49] referred exclusively to textual media (at best containing pictures), the approaches can be applied to other types of media. Figure 35 shows their approaches adapted to image annotation. The top left image shows the regions for which an annotation exist can simply be displayed on the image. The textual comment - if any - is displayed in close proximity to the image, in the example to the right of the image. Annotations can be aggregated to regions where annotations accumulate the same way as with text, an example can be seen in the top right image of Figure 35. The example shows two overlapping aggregates of annotations around the sun in the top left part of the image. Lastly, the list of annotated content and annotations can be built the same way as with textual content with the exception that the part the annotation is referring to is not a passage.
of text but a part of the annotated image. An example can be seen in the bottom image in Figure 35.

The adaption to video annotation assumes that temporal context is used, and the scrub bar of the video is used to create and visualize the points in time and time intervals an annotation is referring to. Examples for this can be seen in Figure 36: The different time intervals are visualized as bars on the scrub bar, and the comment is shown in a list to the right of the video. The first approach can be adapted as it stands to video annotation: The different time intervals and points in time are visualized as bars of different length on the scrub bar. The textual comment - if any - is displayed in proximity of the video, in the images of Figure 36 in the margins right to the video. Aggregates of annotations can be calculated using the context information available and displayed on the scrub bar - a possibility for visualization can be seen in the top right image of Figure 36. The last approach is more complicated compared to text or images because there is a temporal component involved. For visualizing the part of the medium the annotation is referring to, two different possibilities are conceivable: Display a set of stills from the part of the video the annotation is referring to or embed the part of the video the annotation is referring to itself. The latter version is sketched in the bottom image of Figure 36.
5.1 Displaying Annotations

The third approach assumes that passages where annotations accumulate are in some way interesting parts of a medium. Studies which examine this assumption are discussed in the next section. Admittedly, it is only examined for textual media. If those findings can be applied to image and video content is not clear and should be part of future work.

5.1.1 Identifying Interesting Passages

First evidence indicating that annotations accumulate on interesting passages was found by Marshall [46]: She examined physical copies of the same textbook annotated by students. The number of sentences per chapter that every student had annotated was higher than if annotations would have been created randomly. This fact led Marshall to the assumption that there are interesting passages which can be identified by being annotated by a number of students. She discarded the possibility that the annotations are at passages that stand-out visually, e.g., by typographic idiosyncrasies or being headings. The majority of the overlapping annotations were deep within paragraphs. With her study, Marshall created the foundation for further studies in the area of identifying interesting passages of documents using accumulations of annotations. [46]

Shipman et al. [65] try to identify so-called high-value annotations with the goal to make rereading of documents easier. The assump-
76 COLLABORATING ON ANNOTATIONS

tion behind this is that a set of high-value annotations can be used to gain an overview of a document. For their study, they examined the number of citations used in a legal brief which refer to an annotation in a referenced document, i.e., they examined which of the annotations of a user led to a citation. They observed that passages which were marked with more than one mark type, e.g., underlined in conjunction with a comment in the margin, were more likely to be used for citation. They implemented their findings in form of a parser that clusters and weights groups of digitally created freeform annotations. They attached higher scores to annotations the more selective they were, i.e., creating an area around text as opposed to selecting text directly. An interpretative mark, e.g., a note or a comment, got an even higher score. Depending on a pre-defined threshold for the score, clusters were displayed in different colours to indicate different grades of importance. [65]

While Shipman et al. [65] examined only left by a single person in a document, the usage of clustering is similar to the method that would be used for annotations left by different annotators. Attaching different weights to annotation types is certainly a good idea because it is easily conceivable that an accumulation of textual annotations is a better indicator for importance than an accumulation of highlights without accompanying text.

Bradshaw and Light [8] examined overlapping annotations left by biology graduate students. This study was already discussed in my bachelor’s thesis. [43] In their study, they addressed three different research questions [8, p. 209f.]:

- How much overlap in annotated passages exists between different annotators? If there is no overlap, interesting annotations can not be identified by passage consensus.

- How much of the document is annotated? If the majority of a document is annotated, passage recommendation makes no sense because with the whole document being interesting, readers would have to read the whole document none the less.

- How are the annotations distributed throughout the document? If the interesting passages are always found at certain parts of a document, passage recommendation makes no sense because annotation consensus would always point to those passages.

Their study revealed that annotations are distributed over the whole document and that 80% of all annotations are contained in 33% of all sentences. For annotation overlap, they observed “that the majority of annotated content is marked by 50% or more of the readers.” [8, p. 213] The authors conclude that overlapping annotations “may make an effective basis for directing later readers to passages they would find useful.” [8, p. 216] They admit that there may be a difference in
interesting passages identified by annotation consensus if people annotate and read the material for different reasons. [8] On a collaborative learning platform users usually consume material for a common goal and purpose - acquire knowledge to pass the exam.

The approaches and studies presented in this section focused exclusively on textual content, either in form of text books, legal references, or scientific papers. It is conceivable that the described approaches can be applied to lecture notes but that and their applicability to other kinds of media have to be shown through further studies.

5.1.2 Wrapping up Displaying Annotations

This section introduced different approaches for displaying annotations adapted from Marshall et al. [49] There is the possibility to display all annotations directly on the medium, display aggregates of annotations on the medium, or display a list of comments accompanied by their context.

All three approaches make sense for a collaborative learning platform: The first approach can be used in conjunction with means for filtering and ranking annotations. The third approach can be used besides the first or second one to give users the possibility to get an overview of all annotations for a medium, again, coupled with means for filtering and ranking. The second approach assumes that passages where annotations accumulate are interesting passages - for that, studies examining this relationship were introduced.

The next part of this chapter introduces different possibilities for creating a ranking of annotations.

5.2 Ranking Annotations

The Oxford Advanced Learner’s Dictionary defines ranking as “the position of (...) something on a scale that shows how good or important they are in relation to other similar (...) things (...)” [31] Rankings enable a collaborative learning platform to order annotations, so that annotations that are in some way better than other appear first in a list of annotations. That means, ordering annotations is the process of putting annotations in order using some kind of ranking. Filtering is a closely related term: Filtering means to remove annotations that do not satisfy some kind of predicate. This predicate can be formulated using a ranking, e.g., in form of a threshold.

As already discussed, means for ordering and filtering are important aids when working on a collaborative learning platform. If users are overwhelmed by bad or irrelevant annotations, working with the platform becomes nearly impossible. In this section, different rankings are discussed: From a simple ranking, using the creation time, to
rankings using the expertise of the annotation’s author, to rankings using the rating of an annotation.

5.2.1 Creation Time as Ranking

The easiest ranking is the creation time of an annotation. It can be used to present users a chronologically ordered list of annotations. While there are applications where ordering by creation time makes sense, for a collaborative learning platform this order is most likely unsuitable because annotations, as opposed, e.g., to a news article, are not a time-dependent medium.

The desired ranking for a collaborative learning platform should achieve an ordering by quality or relevance of an annotation, and an annotation is not necessarily of higher quality the more recent it is. In the worst case, the best annotation is the annotation that was created first which can be found at the end of the list. But, there are also use cases where chronological ordering is reasonable: Imagine a lecturer who wants to answer the most recent questions about a lecture. Here, a chronological order makes the most sense.

The implementation of ordering by creation time takes next to none implementation effort and can be, additionally to its use cases, used as a baseline in further studies in this area.

5.2.2 Annotator’s Expertise as Ranking

Another possible ranking is the expertise or knowledge of an annotation’s author. This section introduces various approaches to measure and determine a person’s expertise. Expertise and knowledge are not binary which is an important factor for a collaborative learning platform where there is a variety of topics: A person can be an expert for a certain topic but a layman in another one. So, the challenge is to determine the topics of the platform and for each topic the users that are experts in that topic.

Wang et al. [75] propose four categories in which expert finding mechanisms can be classified in: Using self-disclosed information, using self-authored documents, using social network analysis, and using hybrid approaches. They also discuss different approaches and studies for each category, many of them can also be found in the following section.

Expert finding mechanisms using social network analysis often use eigenvector centrality for determining experts. The next part of this chapter gives a brief overview of the two most-used eigenvector centrality measures, PageRank and HITS. After that, different expert finding mechanisms are introduced.
5.2.2.1 Theoretical Foundation

There are expert finding mechanisms that exploit eigenvector centrality to gain an importance or expertise ranking. For the understanding of the following sections, the arguably most important of those measures, PageRank and HITS, are introduced briefly. For the rest of the thesis, the following definitions are used to describe graphs.

Definition 1 (Graph). A graph is a tuple $G = (V, E)$ with $V$ being a set of nodes and $E$ a set of edges with $E \subseteq V \times V$. A directed graph is a graph where each $(u, v) \in E$ is directed from $u$ to $v$.

Definition 2 (Adjacency Matrix). The adjacency matrix of a graph $(V, E)$ is a $|V| \times |V|$ matrix with a cell $a_{ij}$ being 1 if $(i, j) \in E$ and 0 otherwise.

PageRank

PageRank is the reason Google exists: Google was built as a test bed for the ranking algorithm developed by Page et al. [58] The main idea of PageRank is that the importance of websites flows evenly through the graph induced by the link structure of the internet, i.e., two pages (represented as nodes of the graph) are connected by a directed edge $(u,v)$ if website $u$ has a link to website $v$. That conforms to the intuition that a page gives their importance to the pages it is linking to. [58] For example, a website with the fictional importance of 6 containing 3 outgoing links would give an importance of 2 to each of the sites the outgoing links are pointing to.

Formula 1 describes the intuitive explanation from above mathematically: The importance of a website $u$ is determined by the sum of the evenly distributed importance of each websites pointing to it.

$$R(u) = c \sum_{v \in B_u} \frac{R(v)}{N_v}$$  \hspace{1cm} (1)

where

- $R(a)$ : importance rating of website $a$
- $B_a$ : set of all websites containing a link to website $a$
- $N_a$ : number of outgoing links on website $a$

Figure 37: Formula for calculating the importance score of website $u$. Adapted from [58, p. 3].

The dependency in Formula 1 can be modelled for all websites at the same time using a matrix-vector multiplication: In the transposed adjacency matrix $A^T$ of the graph described above, each row conforms to the links pointing to the page represented by the row. Prior
to that, A is made column-stochastic by dividing each entry of a column by the sum of all entries of the respective column. The column vector R contains the importance score for each website. $R = AR$ is the equation to fulfil: The importance scores form an equilibrium on the matrix A. Mathematically speaking, R is an eigenvector of the matrix A for an eigenvalue of 1. [58]

A possible problem is that the graph potentially contains subgraphs that have only incoming edges and otherwise only edges between nodes within the subgraph. This leads to a scenario where importance flows into the subgraph but never leaves the subgraph which leads to an accumulation of importance in the nodes of the subgraph. Page calls such nodes rank sinks. To remove rank sinks from the equation, Page et al. introduce another vector that represents random visits to other websites, i.e., modelling the behaviour of a user not always following links on a site but manually entering a new URL in the address bar of the browser. The resulting equation is $R = (AR + E)$ with $E$ being a column vector. [58]

$$R^{n+1} = (AR^n + E)$$ (2)

Figure 38: Power method for calculating PageRank. Adapted from [58, p. 6].

A method for calculating PageRank, called power method or power iteration, can be seen in Formula 2. The power method is carried out by repeatedly calculating Formula 2 until the difference in results from two consecutive calculations is below some pre-defined threshold. For the matrix used to calculate PageRank, the power method is guaranteed to converge. [58]

Hypertext Induced Topic Search

Another algorithm, which exploits eigenvector centrality, emerged around the same time as PageRank, is Hypertext Induced Topic Search (HITS) by Kleinberg [41]. As opposed to PageRank, HITS calculates two scores, an authority and a hub score. [41]

The authority score defines how relative or definitive a website is with respect to a query. The higher the authority score, the more relevant a website. Hubs are websites “that link to many related [to a topic, author’s remark] authorities.” [41, p. 607] Hence, the hub score should be higher the more links to good authorities exist on the respective website. Between hubs and authorities exists something that Kleinberg calls “mutually reinforcing relationship” [41, p. 611] which means that “a good hub is a page that points to many good authorities (...) [and] a good authority is a page that is pointed to by good many hubs.” [41, p. 611] HITS does not operate on the whole graph
induced by the link structure of the internet but on a subset of nodes relevant to a search query and the links between them: First, a set of $t$ relevant websites, called root set, is determined and extended by nodes that point to websites in this set and websites that are pointed to from within the root set. This extended set, called base set, is the base for the HITS algorithm. [41]

$$A^{n+1} = L^T H^n \quad (3) \quad H^{n+1} = LA^{n+1} \quad (4)$$

where

- $L$: adjacency matrix of the graph
- $A^n$: column vector containing authority scores after iteration $n$
- $H^n$: column vector containing hub scores after iteration $n$

Figure 39: Formulas for calculating authority and hub score using HITS. Adapted from [41, p. 612].

The formulas for resolving the circular dependency described above are shown in Formula 3 and 4. The authority score is calculated by using the hub score from the previous iteration, while the hub score is calculated by using the new authority score from the current iteration. That means the authority score always has to be calculated before the hub score. As with PageRank, for calculating hub- and authority scores the power method is used. Further discussion of possible applications and convergence criteria can be found in Kleinberg’s publication. [41]

**Expert Finding using Self-disclosed Information**

As the name already suggests, expert finding mechanism based on self-disclosed information use information given by the users themselves to find experts in a certain area. Wang et al. [75] criticize this expert finding method for being time-consuming for users and the unlikelihood that users maintain their profile, i.e., add new skills and update existing skills. [75]

An example where self-disclosed information is used to find experts is XING\(^1\) where users can specify qualifications and skills they are offering to potential employers. This can be seen as a form of expert finding because recruiters use this information to search experts (i.e., new employees) with a particular set of skills.

---

\(^1\) [https://www.xing.com](https://www.xing.com)
LinkedIn\(^2\) and ResearchGate\(^3\) expand on this approach: Users can specify areas they have expertise in, but additionally, other users have to possibility to endorse people on those skills. Endorsement results in some kind of validation of the skills people add to their profiles.

On a collaborative learning platform, it is conceivable to give users the possibility to state areas they are experts in. Preferably, these areas come from a controlled vocabulary to prevent that users use different terms to describe the same knowledge area.

A collaborative learning platform has access to textual content authored by the user, e.g., created annotations or a seminar paper created using the platform. This opens the way for another kind of expert finding - using documents authored by the user.

**Expert Finding using Authored Documents**

Wang et al. \([75]\) state that documents “written or reviewed by an expert candidate” \([75,\ p.\ 1443]\) can be used to find experts. In the context of a collaborative learning platform, the source of written content by possible expert candidates would be mainly annotations, but other sources, such as a seminar paper created on the platform, are conceivable as well.

Techniques using authored documents to identify experts share a similar structure: The topic or query for which an expert shall be found is represented as a set of words, and the relevance to documents in the collection is calculated using similarity measures. The relevance scores are somehow aggregated to calculate a ranking for experts. According to Liu et al. \([42]\), the intuition is that authors who wrote many documents in a knowledge area are likely experts in the respective area.

Balog et al. \([4]\) use the term *knowledge area* to describe an area in which experts shall be determined. Their idea is to calculate the conditional probability \(q(\text{ca}|\text{q})\), in other words, given a query \(q\), the probability of candidate \(ca\) being an expert, with the candidates with the highest probability being experts. To determine this probability, they use Bayes’ Theorem and generative probabilistic language models to calculate the missing variables required to solve the Bayes’ Theorem. They introduce different ways to calculate the missing variables that focus either on the candidate or the document. Best results were achieved when they calculated the relevance between documents and knowledge areas first and later used a user-centred association between users and documents to find experts. \([4]\)

Building upon the results from this study, Balog and de Rijke \([5]\) tackled another related task: Expert profiling. As opposed to expert finding, which is the task of finding an expert with respect to a knowledge area or search query, expert profiling aims to build a profile

---

\(^2\) https://www.linkedin.com

\(^3\) https://www.researchgate.net
for candidates containing knowledge areas they are an expert in. To achieve this, they chose a similar approach as Balog et al. [4] took in the previous study: They take a list of pre-defined knowledge areas and determine a value that expresses the knowledge of each candidate in each knowledge area. For determining the score, they propose two different methods: Using a generative language model or using keyword similarity. [5]

The first method determines a set of documents relevant to a knowledge area that are associated with the current expert candidate and defines the score as the sum of all relevance scores between document and knowledge area. The intuition is that an author that wrote many relevant documents in a knowledge area is most likely an expert. Relevance is determined using a generative language model. The other method extracts the top 20 keywords from documents of the knowledge area as well as the top 20 keywords from documents of the knowledge area associated with the user. The score is defined as the percentage of all keywords of the current knowledge area that also occur as keywords in the documents from that knowledge area associated with the respective user. The intuition behind this method is that a user who uses many keywords of a knowledge area is most likely an expert in that knowledge area. Their results show that the first method is slightly better than the second one and that filtering can be used to further enhance the results. [5]

Liu et al. [42] describe another method for expert finding using authored documents: As opposed to Balog and de Rijke [5], they do not rely on a list of pre-defined knowledge domains but are using Latent Dirichlet Allocation (LDA) to determine knowledge domains. In this context, LDA is used for clustering with input being the documents $D$ and the desired number of clusters $T$ and output being $T$ clusters each containing words from $D$ that are most likely related to each other. After that step, the relevance between the documents and the knowledge domains represented by the clusters is determined. [42]

For this task, Liu et al. [42] examined three possibilities: Using the Vector Space Model, a probabilistic language model, similar to Balog et al. [4], or LDA. After that, there is a matrix of documents and knowledge domains with each cell representing how relevant a document is for a knowledge domain. In order to determine experts for each knowledge domain, this matrix has to be associated with users which is achieved by using a user-document association function. The authors propose two different formulations for this function, a direct one and an indirect one. The direct one is $1$ if the user wrote the document and $0$ otherwise. The indirect version is $1$ if the user wrote the document, a value between $0$ and $1$ for different degrees of association, and $0$ if there is no association between them. [42]

They evaluated all possible combinations of relevance measures and association functions described above against each other: Their
results show that the best results were achieved when using LDA for calculating the relevance score in combination with the direct association function. Even better results were achieved when the documents were filtered before, and only the top part of the relevant documents was taken into consideration for expert finding. [42]

On a collaborative learning platform, there is not only self-authored content but also a social network that can be used for expert finding. The next section introduces various approaches that try to exploit the social structure of a platform for finding experts.

**Expert Finding using Social Network Analysis**

Expert finding using social network analysis tries to exploit the structure of a social network for finding experts. [75] Most of the approaches exploit eigenvector centrality using one of the two methods described in Section 5.2.2.1.

Zhang et al. [79] study different measures for identifying experts on the example of a forum about the programming language Java. There are generally two types of actions in a forum: Asking questions and answering questions. The authors model these actions in form of a user-user graph where each node corresponds to a user and each directed edge \((u, v)\) representing that user \(u\) answered a question from user \(v\). The intuition of this model is that a user can only answer a question if he is of higher expertise than the questioner, and the questioner’s expertise should be distributed evenly among all people (i.e., all answerers) that provided an answer. [79] For this graph they proposed three different measures for expertise: [79, p. 224f]

- **Z-score** assumes that a person with greater expertise answers more questions than asks questions. Zhang et al. [79] calculate the z-score as the difference between the user and a (fictional) user that randomly posts questions and answers. A modification of z-score called z-degree only considers unique users, e.g., a user that always helps the same user has lower expertise than a user that helps many different users.

- **ExpertiseRank** uses the PageRank algorithm on the unweighted adjacency matrix of the user-user graph.

- **HITS authority** uses the authority score of HITS on the unweighted adjacency matrix of the user-user graph.

For evaluation of the measures, the authors first assessed the expertise of a set of users from the dataset on a scale from 1 to 5, with 5 being the highest level of expertise. Their evaluation showed that “z-score-based ranks tend to produce slightly better results than other methods.” [79, p. 226]

Zhang et al. [80] applied their findings from the study described above to develop an application, called QuME, that matches experts
to questions on the examined Java forum. To determine a suitable user, they use the measures for expertise described above coupled with other information, such as the recency of the question, the number of already existing answers, and the similarity of the possible answerer’s profile to the question. They assumed that users are more likely to answer questions that are only slightly below their expertise than questions that are greatly below their expertise, so the difference between questioner’s and possible answerer’s expertise was used as a factor as well. They implemented a prototype but did not perform an evaluation, so that while the intuition behind the algorithm seems reasonable, its real-life performance is unknown.\[80\]

Jurczyk and Agichtein\[37\] use the authority score of HITS for finding experts: They evaluated their approach on two versions of the user-user graph described above. One version collapsed duplicate edges, i.e., edges where a user answered two or more questions of another user were collapsed into a single (unweighted) edge, and another version where the edges were not collapsed. The latter version is tantamount to a single weighted edge with weight being the number of times the user answered one of the other user’s questions. They evaluated their algorithm on a subset of three categories from Yahoo! Answers. Their evaluation shows that the version with unweighted edges had poorer performance in the expert finding task. They established their gold standard for experts by looking at data that can be gathered from the data set, e.g., up- and downvotes or rating of an answer. For most question categories, their algorithm achieved good performance in finding the top experts.\[37\]

The algorithms introduced in this section were used in situations where the area of expertise was already set: Java in case of Zhang et al.\[79, 80\] and the three categories from Yahoo! Answers in case of Jurczyk and Agichtein\[37\]. When the dataset is bigger and the areas of expertise are not pre-defined, hybrid approaches are required.

**Expert Finding using Hybrid Approaches**

As already suggested by the name, hybrid approaches to expert finding combine information from different sources to calculate a ranking for the expertise of a person.\[75\]

ExpertRank is an expert finding algorithm proposed by Wang et al.\[75\] that combines a *relevance* and an *authority score* to an *expert score*. The relevance score describes the similarity between the query and a single document containing all textual content created by the expert candidate. As a measure for similarity, they use the Vector Space Model but state that other similarity measures are conceivable as well. An expert candidate’s authority score is determined using the PageRank algorithm on a modified user-user graph. As opposed to the user-user graph used for example by Zhang et al.\[79\], they
weight edges differently: Edges that exist in both directions (even indirectly, with more than one node on the path between two nodes) are weighted lower than edges that only exist in a single direction. For the combination of the two scores, they propose three possibilities: Linear combination, which is a fancy term for a weighted sum, cascade ranking, which ranks users using the relevance score first and then selects experts from the top of the ranked list, and a scaling strategy. [75]

Evaluation of their algorithm took place on the Microsoft Office forums where helpful users are tagged as most valued professionals (MVP) which were taken as the gold standard for experts. They achieved the best results using the cascade combination strategy which outperformed both expert finding using only the relevance or authority scores. [75]

Pal and Counts [59] propose an algorithm that can be used to identify “topical authorities in microblogs” [59, p. 45], i.e., people that are experts of a certain topic in microblogs, such as Twitter. Their algorithm takes a set of filtered annotations about a certain topic as a foundation which, e.g., can be obtained by a simple keyword search for the topic. After that, the set of tweets is divided into two clusters of users: A cluster where authorities are most likely to be found, and a cluster where most likely no authorities are to be found. For effective clustering, they define numerous features on the set of tweets, e.g., a self-similarity score, which measures the similarity between a user’s tweets, or how much of the tweets for a certain topic is provided by the examined user. The full list of features can be found in the publication. In the end, the users in the cluster most likely containing authorities are ranked using a Gaussian ranking algorithm. They evaluate their approach against text and link-based approaches. Their results show that their approach yields the best results, with the link-based one coming in second. [59]

Agichtein et al. [1] do not directly identify experts but identify high quality content in social media. They use a dataset from Yahoo! Answers for the evaluation of their algorithm. Three types of information are collected from the dataset and used to train a machine learning algorithm with the goal to obtain a classifier that is able to automatically decide if a piece of content is of high-quality. The first category of information is intrinsic content quality which includes measures for typos, syntactic and semantic complexity, and grammaticality. User relationships is an authority score calculated either via the PageRank or HITS algorithm on a user-user graph. [1]

The last kind of information are usage statistics which includes data such as the “number of clicks on an item.” [1, p. 186]. The best performance achieved a classifier built using stochastic gradient boosted trees which achieved “a high level of accuracy on the question and answer quality classification part.” [1, p. 193]
There are other hybrid expert finding mechanisms, such as Xpert-finder [67] which uses the similarity between all of a sender’s emails and topic descriptions in conjunction with graph analysis. Campbell et al. [11] use HITS’ authority score to identify experts using email content and graph analysis on a pre-filtered set of emails. Dom et al. [23] use various measures for expertise on a pre-filtered set of emails.

5.2.3 Rating as Ranking

As already discussed in Section 4.1.5.3, a rating is a value that expresses the subjective degree of appreciation for something. There are two different manifestations of rating: Implicit and explicit rating. Explicit rating is every kind of rating where the user consciously creates the rating, e.g., by pressing on the thumbs-up button on Facebook. Implicit ratings are inferred from interactions of users with the system. Examples for implicit rating are the number of clicks on an item or recurring visits. [62] Recurring visits and their frequency were examined and studied in detail by Jan et al. [36] for usage as implicit rating.

This section discusses approaches for determining a ranking from both explicit and implicit ratings and their possible sources on a collaborative learning platform.

By Explicit Rating

The simplest approach for accumulating explicit user ratings to a value that can be used as a ranking is the arithmetic mean. There are several downsides to this measure: It is possible that two items have the same arithmetic mean even if one item was rated by two people, while the other one was rated by two thousand people. In this case, the latter situation can easily be seen as the more significant and therefore should be ranked higher. [62]

Miller [52] discusses different methods for determining rankings in a blog entry. According to him, average rating (i.e., arithmetic mean as described above) can be used if many individual ratings are available. Another ranking measure, the difference between positive and negative ratings, is dismissed as impractical because its is possible that an item with a greater percentage of positive ratings is ranked below an item with a lesser percentage of positive ratings using this measure. [52]

Pohl [62] proposes the Bayesian Average aggregation as a possible way to aggregate the ratings into a ranking. Miller [52] proposes the lower bound of the Wilson score confidence interval for a Bernoulli parameter for aggregation. Discussion of the performance of possible rating aggregation methods is outside the scope of this thesis. An comparison between the two methods has been conducted by Pohl.
and can be found in [62, p. 96] where the Wilson score confidence interval seems to produce the more intuitive results.

On a collaborative learning platform, rating and liking annotations and other content are integral concepts. The ratings collected there can be used to calculate a ranking. But, as stated by Pohl, one problem with explicit ratings is “data sparsity” [62, p. 93], i.e., the lack of rating data. One way to alleviate data sparsity is the use of implicit ratings which are discussed in the next section.

**By Implicit Rating**

The ways of aggregating implicit ratings to a ranking are dependent on the type of implicit feedback used to generate the rating. One possibility would be to take the evidence provided by implicit feedback and transform it into something where one of the methods discussed in the previous section can be used.

Jan et al. [36] calculate a *confidence score* for annotations using implicit ratings to introduce two of filters into a collaborative learning environment. The first filter, called *high-grade filter*, filters out annotations whose confidence score is below a pre-defined threshold. The second filter, called *master filter*, only shows annotations created by one of the five *master annotators*, i.e., the five users with the highest average annotation confidence score. The annotation confidence score is determined using the time the annotation was last visited and the average time between successive visits of an annotation. Intuition behind this choice is that a good annotation is visited more often than a bad annotation. [36] For evaluation, an expert graded the annotations, and the comparison of those grades with the scores produced by the algorithm showed “good agreement between the annotation filter and [the annotation grades] of a senior chemistry teacher.” [36, p. 90]

For future work, Jan et al. propose further implicit feedback that can potentially be used for ranking, e.g., “the length of engagement” [36, p. 90], i.e., the duration a user dwells on a piece of content. While Jan et al. used their confidence scores for filtering, they can as well be used for ranking.

Sources of implicit ratings on a collaborative learning platform are numerous, and besides the ones discussed by Jan et al., there are different sources of implicit feedback discussed in Section 3.5 that can be considered to be used as implicit rating. On a collaborative learning platform, not only the annotations, but the whole system can be seen as a source for implicit ratings: A user that is subscribed by many other users creates potentially high-quality content, a lecture which is often annotated is potentially a high-quality lecture, and a lecturer with many high-quality lectures is potentially a high-quality lecturer.
5.2.4 Filtering Annotations

As already discussed in Section 5.2, filtering and ranking are two similar concepts: Ranking can be the foundation for filtering, e.g., by defining a filtering predicate using a ranking score as a threshold. Filtering can be done as well using other attributes of an item, such as the creation time or the creator.

Filters can be pre-defined, e.g., filter out all annotations with a quality score (a ranking) below 0.5 or only show the annotations with the highest quality score. Filters can as well be user-defined, e.g., show only annotations by a certain group of users or created within a specified time interval. Another distinction is between simple and complex filters: A simple filter allows users to filter out content that has a certain attribute, while a complex filter allows users to filter out content that lies outside of a certain range. An example for a simple filter is filtering by annotation type, and an example for a complex filter is filtering by a time interval, i.e., only display annotations that were created within a certain span of time.

Many types of filters are conceivable, and it has to be studied which of them are really required for a collaborative learning platform. Simply implementing every kind of filter is no solution because it would introduce unnecessary elements to the user interface and potentially overwhelm and confuse users. The right set of filters is most likely dependent on the situation and task annotation is used for and has to be found in real-life studies.

5.2.5 Wrapping up Ordering Annotations

In this section, various approaches for calculating ranking for annotations were discussed: Simply using the creation time as ranking, using the expertise of an annotation’s creator, or by ranking using ratings, either explicit or implicit.

For determining the expertise of an annotation’s creator, algorithms for expert finding where introduced which used different approaches: Either using information entered by the users themselves, documents authored by users, the social network structure, or hybrid approaches which combine two or more approaches.

Rating and ranking were discussed and how ratings can be aggregated into a ranking. Implicit feedback which can possibly be used for determining ranking was discussed as well. Filtering, as another mechanism to reduce the information flood, was discussed in the final section.

Other approaches for ranking include using the reputation of annotators or the relevance of annotations. A discussion about reputation can be found in [62, p. 103].
5.3 WRAPPING UP COLLABORATING ON ANNOTATION

This chapter discussed means to reduce the cognitive load on users of a collaborative learning platform: Discussed were both different concepts for displaying annotations and approaches for ranking, ordering, and filtering annotations.

For display, concepts introduced by Marshall et al. [49] were introduced for text and adapted for image and video annotation. The next part discussed various rankings that can be used either for ordering or filtering annotations.

The next chapter discusses the actual implementation of the annotation framework taking into account lessons drawn from the previous sections.
This chapter introduces Annoto a framework for annotations built using the information gathered in the previous chapters. It is built to support a great variety of media, contexts, and annotation contents. Focus was placed on the re-usability of functionality between media types and a structure that allows easy extensibility. Annoto is implemented in JavaScript using jQuery\(^1\), mustache.js\(^2\), and Ink\(^3\). The following description of the framework makes use of concepts that can not be found in JavaScript, such as interfaces and abstract classes. While those concepts do not exist in JavaScript, the software components behave and have to be used as if they were instances of said concepts.

![Annotation Context Content](image)

Figure 40: Simple model of an annotation.

The annotation model that lies at the heart of Annoto is shown in Figure 40. It describes an annotation as a composition of content and context. Context is all information required to identify the part of a medium an annotation is referring to. As the name already suggests, content is the content of an annotation, e.g., a textual remark. The following sections discuss the implementation of the model in the framework.

### 6.1 ANNOTO

Annoto is a framework which can be used to created so-called annotation components which are web applications that allow the annotation of media of a certain type. The two main pillars of the framework are the sharing of media-independent functionality and the easy creation of new types of annotations.

Media-independent functionality is contained in a class that is superclass of all annotation components, called *Annotator*. The Anno-
tator deals with tasks such as setting up the workspace, which includes creating the buttons for changing colours or line width, offering comment functionality, or means for sorting and filtering annotations. Each specialization of the Annotator is responsible for a certain type of media and is called Media Annotator (e.g., Image Annotator for the image annotation component). A Media Annotator displays the medium, offers control of the medium, and exposes attributes of the medium to other parts of the component.

The easy creation of new types of annotations is enabled by encapsulating the process of creating the context of an annotation and the content of an annotation within two separate classes that can be arbitrarily combined. The process of creating the context of an annotation is encapsulated in a Selection, while the process of creating the content of an annotation is encapsulated in an Annotation Type.

The following chapter introduces the structure of an annotation component and explains the four main actors in an annotation component and their interaction in detail. This chapter is by no means a complete documentation and only serves as an introduction to the main concepts behind Annoto.

6.1.1 Structure of an Annotation Component

Figure 41 shows the structure of an annotation component: As already explained, a Media Annotator inherits all media-independent features (and a few more which are explained later) from the Annotator. Media-independent parts are shaded in grey in Figure 41.

To allow annotation, at least one Selection has to be defined for a Media Annotator. After a Selection has finished creating a context, there is context for an annotation, but no content. To allow attaching content to a context to create an annotation, Annotation Types are used.

6.1.2 Annotator

Most of the functionality offered by the Annotator can be seen in Figure 42. For illustration purposes, an implementation of an image annotation component is used.

Annotations are represented at two different places: Directly on the medium or a surrogate element (e.g., the scrub bar of a video), which is used to visualize the context of an annotation, and additionally in a list of annotations where the content of the annotation and optionally a textual representation (e.g., for time interval) of the context is shown. In the example, the polygonal area is the representation of the context, and the item in the annotation list represents the content, in this case a textual comment. The buttons for deleting an item (button with
cross icon) and commenting on an annotation (button with speech bubble icon) are provided by the Annotator as well.

The current implementation does not allow editing content or context of annotations. The reason for this decision is that editing one part of an annotation could change the discussion context for existing comments and could be abused to put commenter in a negative light. Regardless of those concerns, functionality for editing annotations could easily be implemented.

The buttons for changing colour and line width, which can be seen in the top of Figure 42, are provided directly by the Annotator as well because those properties are used by different types of media.

The toolbar, which can be seen between the colour and line width buttons in Figure 42, is automatically populated with the Selections that are defined for an Media Annotator. In this case, four different selections are defined.

The checkboxes and select field in the top right corner are also supplied by the Annotator: Currently, it is possible to filter Selections out (e.g., do not show polygonal selections) and order by pre-defined order values. New order values can be added easily via a configuration file, while adding new means for filtering is more complicated.
because there are different forms of filtering predicates that have to be implemented.

The features described here can be implemented without knowledge of the type of the annotated medium. Media-dependent functionality is implemented in a Media Annotator.

### 6.1.3 Media Annotator

A Media Annotator is a subclass of the Annotator and has four main tasks: Displaying the medium, providing control of the medium and access to context information and transformation, and provide annotation containers.

Using the image annotation component in Figure 42 as an example, displaying the medium would be displaying an image: In this case, this is done by drawing the image data onto a HTML5 canvas element. The second task, providing control of the medium, means to give users the possibility to change the state of the medium, e.g., by zooming in on an interesting spot or panning the image. The example enables this by offering a pointer tool (first button in the menu of Selections) which allows to pan the image by dragging and zoom in and out using the mouse wheel.

Context information is everything that is required to create context for an annotation, e.g., the current page when creating an annotation for a PDF document. Context transformations are functions that transform coordinates from a user’s current viewport into system-independent coordinates which can be used to display annotations on any system. An example is a function that takes coordinates where a mouse click has occurred on the current representation of an image (i.e., can be translated and zoomed) and transforms those coordinates in coordinates independent from translation and zoom.

---

4 Published under CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0). Changes from source material: Annotations. Available online: https://en.wikipedia.org/wiki/File:Eyjafjallajökull.jpeg
Annotation elements are HTML elements where annotations or prompts for creating annotation content are rendered in. Currently, there are two kinds of annotation elements: A HTML div element and an HTML5 canvas element. The div element is used to display dialogue windows for creating content of an annotation (e.g., a dialogue window for entering a textual comment) or the indicator for the existence of a sticky note. A canvas annotation element is used to display other kinds of annotations, such as highlights or the polygonal region shown in Figure 42.

Figure 43 shows the arrangement of the annotations elements for the example of the image annotation component: The bottom most layer is the rendered image itself, above that is a div element, and on top is a canvas element.

A Media Annotator has at least one Selection. Selections are responsible for context creation and are discussed in the following section.

6.1.4 Selection

A Selection encapsulates the process of creating the context of an annotation. An example is the creation of a polygonal shaped region which can be seen in Figure 44: A polygonal shaped region is created by clicking on the desired position on the canvas which creates the first corner of the polygon (top left image). Further corners are added by the clicking at the desired positions on the canvas (top right image). To finish the creation of the region, the first corner (bottom left image) has to be clicked on again. After that, the menu for creating content is shown, which is explained in the next section.
A Selection defines a map that maps JavaScript events\(^6\) to function references. The functions the references are pointing to are to be called if the event occurs and the Selection is currently active. Those functions are defined within the respective Selection and handle the creation of the context, e.g., drawing the current progress onto the annotation canvas. Another task of a Selection is the definition of the context structure, i.e., the structure of the context saved in the database. For the polygonal area mentioned above, the context structure is a list of points.

A Selection also has to implement functions to create both representations of an annotation, i.e., the context representation on the medium or surrogate element and the display in the annotation list. Functions that check if an annotation is currently visible with respect to the current state of the viewport (i.e., with respect to translation and zoom with at the example of image annotation) and functions that check if an annotation is on a certain position (to enable selection of annotations) must as well be implemented in a Selection.

A Selection is media-dependent because it relies on context attributes exposed by the Media Annotator for creating a context.

The last image in Figure 44 shows the menu of all Annotation Types loaded for that Selection. A Selection has at least on Annotation Type, and those are displayed after a Selection has been finished, i.e., the context for an annotation has been successfully created. In the example in Figure 44, two Annotation Types are defined, the Empty

---

\(^5\) Published under CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0). Changes from source material: Annotations. Available online: https://en.wikipedia.org/wiki/File:Eyjafjallajökull.jpeg

\(^6\) https://developer.mozilla.org/en-US/docs/Web/Events
Annotation (the check symbol), which creates an annotation without content, and the textual annotation (the document symbol), which creates an annotation with a textual comment. The cross aborts the creation of an annotation. Annotation Types are explained in the next section in detail.

6.1.5 Annotation Type

An Annotation Type encapsulates the process of creating the content of an annotation. An example for an Annotation Type is a textual annotation. A simple implementation can be seen in Figure 45.

Figure 45: Dialogue window shown when creating textual content for an annotation. Annotated medium: Eyjafjallajökull by Andreas Tille.

A list of all defined Annotation Types for a Selection is shown after context was created successfully. Choosing one of the Annotation Types gives control to the respective Annotation Type.

An Annotation Type offers functionality for initiating an interaction that leads to the creation of content, e.g., by displaying a dialogue window, as show in Figure 45. After that, the entered data has to be validated, the annotation created, and lastly, added to the Annotation List whereupon the annotation is displayed on the medium and in the list of annotations to the right of the medium.

Annotation Types are media-independent, i.e., they can be used in conjunction with any kind of Selection. This section implicitly introduced the term Annotation List which is a concept used by Annoto to store annotations.

---

7 Published under CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0). Changes from source material: Annotations. Available online: https://en.wikipedia.org/wiki/File:Eyjafjallajökull.jpeg
6.1.6 Storage

Annotations are saved in an instance of the class Annotation List. An Annotation List offers functions for creating, reading, updating, and deleting annotations. The Annotation List represents the only access point to annotations. An illustration of the concept behind the Annotation List can be seen in Figure 46.

![Diagram: AnnotationList and Storages in Annoto](image)

Figure 46: AnnotationList and Storages in Annoto: An Annotator has an AnnotationList where all annotations are stored using the defined Storage. To change the underlying data access layer the Storage in the AnnotationList has to be switched out against another implementation of the interface Storage.

An Annotation List has a Storage which is an abstraction of the underlying data access layer. Examples for data access layers are HTML5’s Web Storage or a REST API. Currently, the only implementation of Storage is RAM Storage which uses a JavaScript array for storing annotations. Storages can be switched out by changing a single line of code.

6.1.7 Content Provider

Content Provider is a concept that is exclusively used for video and audio annotation where the problem exists that content is not always available for storage on one’s own servers. That means that there has to be the possibility to include embedded players from video hosting providers, such as YouTube or Vimeo, but also the possibility to include content hosted on a own server.

Instead of implementing a unique Media Annotator for each source of video and audio content, the concept of a Content Provider was introduced. A Content Provider encapsulates the process of displaying video or audio content and offers functions for starting and stopping playback and jumping to a certain position in the medium. It also sends the current playback progress to the Media Annotator via an event.

---

8 https://developer.mozilla.org/en-US/docs/Web/API/Web_Storage_API
6.1.8 Wrapping up Annoto

The four main concepts of Annoto, Annotator, Media Annotator, Selections, and Annotation Types were discussed in this section. Annoto allows sharing of media-independent features between all annotation components and offers a single point for changing those functionality by containing media-independent features within the Annotator.

An annotation consists of context and content - this separation is represented through Selections and Annotation Types. Selections are media-dependent and implement an interaction that creates a context. Annotation Types are media-independent and implement an interaction that creates content for a context.

Storage is a concept that allows to change the underlying data access layer without the need to modify the system elsewhere. Content Provider are introduced so that content from different sources can be used without the need of implementing a custom Media Annotator for each different source of video and audio content.

Annoto should allow the implementation of most concepts discussed in the previous chapters, especially Chapter 4. Implementing all the discussed concepts would have far exceeded the scope of this thesis, therefore, only a few concepts are implemented as proof-of-concept. The implemented annotation components are introduced in the next section.

6.2 ANNOTATION COMPONENTS

Using Annoto, three annotation components were implemented as a proof-of-concept: One for PDF annotation, one for image annotation, and one for video and audio annotation. Not every concept discussed in Chapter 4 is implemented, instead the annotation components should be seen as proof-of-concept for the framework and used as starting points for further development.

This section introduces the implemented annotation components and the Selections and Annotation Types that were implemented for them.

6.2.1 Annotation Types

Two different Annotation Types were implemented: A textual annotation and an empty annotation. Annotation Types are media-independent and are used in all of the annotation components described in the following.

A textual annotation is an annotation consisting of a textual comment and optionally tags. The current implementation can be seen in Figure 45: A text area for the textual comment and a text field for tags.
An empty annotation is an annotation without content which can, e.g., be used to create an annotation where the focus lies on the context or create a reminder for a passage for which the actual content is to be created at a later point in time.

There are many other Annotation Types conceivable, e.g., a correction annotation, which allows the selection of a score from a predefined range, or a referencing annotation, which allows the user to create a hypertextual reference to another annotation or medium.

6.2.2 PDF Annotation

The PDF annotation component was implemented using PDF.js as library for rendering the PDF document. PDF.js renders the document on a HTML5 canvas element and allows the selection of text by putting span elements on top of the canvas which create the impression that users interact with the text. Figures 47 and 48 show the implementation of the PDF annotation component. The only custom workspace element is the pagination which can be seen at the bottom of both figures. Everything else is provided by the Annotator.

Figure 47: PDF annotation component on a document consisting mainly of graphical content: Sticky note in the top left image of the slide and a freeform drawing showing two circles and an arrow in the middle of the slide.

For the PDF annotation component, three Selections were implemented: A sticky note in the style of Acrobat Reader, highlighting of arbitrary passages of text in one of four colours, and freeform drawing on the document with the possibility to create drawings consisting of more than one stroke with different colours and line widths.

9 https://github.com/mozilla/pdf.js
With freeform drawings, the actual drawing is seen as the context of the annotation - this is, for one thing, easier to implement using the framework, and for another thing, allows to attach additional content to a freeform drawing.

Figure 47 shows a sticky note in the top left image of the slide and a freeform annotation consisting of two circles and an arrow between them. Both annotations have textual content which can be seen in the list to the right of the document. Annotations can be selected from the document or the list which leads to them being highlighted in both representations.

Figure 48 shows examples for highlights: Highlights can be done in one of four different colours for arbitrary connected passages of text.

### 6.2.3 Image Annotation

Image annotation draws the image onto an HTML5 canvas element. Using a canvas element enables the easy panning and zooming of the image. Figure 49 shows the image annotation component. For this component, no custom workspace elements were introduced.

For the image annotation component, three Selections where implemented: A sticky note, which behaves the same as with the PDF annotation component, a rectangular selection, and a polygonal selection, which was already shown in Figure 44. All of those Selections can be seen in Figure 49.
Figure 49: Overview of the image annotation component: A sticky note in the left part of the image, a polygonal selection in the right part of the image, and a rectangular selection in the middle part of the image. Annotated medium: Eyjafjallajökull by Andreas Tille.  

6.2.4 Video and Audio Annotation

Video and audio annotation are implemented in a single component - this is made possible by the similarity between the two media types and the choice to support only temporal context. Currently, Content Provider for three different content sources are implemented: YouTube, HTML5’s video tag, and HTML5’s audio tag. Figure 50 shows video annotation of a YouTube video; the only difference to other Content Provider would be the player in the middle of the image.

Video and audio annotation is not done on the medium itself but on a surrogate element which is the scrub bar below the video. Building an intuitive scrub bar is not an easy task: Most of the platforms examined in Chapter 4 did not allow annotation using the scrub bar and opted instead to an approach where the context was defined via text fields or by the current playback progress.

The scrub bar is shown in detail in Figure 51. It consists of three parts: The annotation creation part above the progress bar, the progress bar in the middle, and the annotation display part below the progress bar. The main idea is that annotations are created on the top part and are displayed on the bottom part. Rationale behind the decision to separate annotation creation and display is to give the user an unob-

---

10 Published under CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0). Changes from source material: Annotations. Available online: https://en.wikipedia.org/wiki/File:Eyjafjallajökull.jpeg
6.2 Annotation Components

Figure 50: Overview of the video and audio annotation component: Two time intervals and a point in time overlapping with the second interval. Annotated medium: Taylor Swift - Blank Space.

Figure 51: Implementation of the scrub bar.

structured space to create annotations, so that context bounds can be clearly recognized.

Currently, there are two Selections defined for the video and audio annotation component: A time interval and a point in time. The creation of a time interval can be seen in Figure 51: The annotator drags the desired interval on the top part of the scrub bar. After that, the menu for choosing an Annotation Type is displayed. A point in time is created by clicking on the desired position in the top part of the scrub bar. After an annotation was successfully created, its context representation is transferred to the bottom part of the scrub bar, which can be seen in Figure 50. For both Selections, a colour can be chosen from the known list of four colours. Annotations can overlap which can be seen in the second interval in the same figure: Here, a point in time is placed within a time interval.

When playing a video back, the currently active annotation is highlighted in the list to the right of the video as well as in the bottom part of the scrub bar. An annotation is active if the current progress of the video is contained within a time interval or within five seconds of a

Available online: https://www.youtube.com/watch?v=e-0RhEE9VvG
point in time. If more than one annotation is currently active, the first one in the list of annotations is highlighted. In a later version, this should be extended to highlight all annotations that are currently active. Annotations can as well be selected from the list of annotations to the right of the video which starts playback at the start of the selected annotation.

6.2.5 Wrapping up Annotation Components

In this section, the annotation components implemented as part of the thesis were discussed: The PDF component uses a library for PDF rendering that allows the interaction with the textual content of the document which is an important requirement for text annotation. The image annotation component allows to zoom and translate the image and even allows users to annotate beyond the borders of the image. Video and audio annotation was implemented using a temporal context and a custom built scrub bar which separates the area where the annotations are created from the area where the annotations are displayed.

The current implementation of the components should be strictly seen as proof-of-concept. It shows that Annoto can be used to implement the majority of annotations described in the previous chapters, and the currently implemented Selections and Annotation Types can be used as a starting point for further development.

6.3 Wrapping up Building the Framework

Annoto is a framework that allows the easy implementation of different types of Selections and Annotation Types. Building a framework was a necessary task to allow the easy extensibility and the sharing of functionality between different types of media. This structure allows to easily create applications that can be used for studies and experiments in the different areas of annotations.

The framework is not finished: There are components that could be made more generic, and Selections could potentially be made media-independent by defining interfaces for accessing context attributes. The same goes for the implemented annotation components: There is still much work to do.

The next chapter wraps this thesis up - it summarizes the findings and gives food for thought for further studies and extensions to the annotation framework.
The goal of this thesis was the development of an annotation framework that allows an easy implementation of annotation components for a variety of media. The framework is intended to enable the fast creation of new annotation settings for further study and research in the area of annotations. The framework, called Annoto, is to be used for a collaborative learning platform that is currently in development. A collaborative learning platform has to offer means for annotation, as annotations play an important role in learning, both in private and collaborative annotation settings which has been shown in different studies.

This chapter summarizes the work done and the findings of the thesis, gives a view in the future of the framework, and proposes future work.

SUMMARY

To build a framework that can be used in a variety of annotation settings and for creating different types of annotations, different forms, functions, and characteristics of annotations were thoroughly examined: Both theoretically, by looking at research in the area, and practically, by looking at annotation means offered by different existing annotation platforms.

Different classifications for annotations were found in the theoretical analysis. The distinction between referencing, textual, and graphical annotations is important because it provides a starting point for the development of different annotation types. A distinction between lecture and reading annotations is particularly interesting in the context of a collaborative learning platform because it suggests that there is a difference between annotations that were created while listening to a lecture and annotations that were created outside of a lecture, e.g., while studying the lecture material. Private and public annotations are relevant when annotation is used in a collaborative setting: A study by Marshall and Brush [48] came to the result that annotation practices differ depending on the intended readership of an annotation.

The results from the theoretical examination were used to create a classification framework for annotation platforms. This framework was used to spot the point at which enough annotation platforms were examined in a benchmarking of annotation platforms. Conversely, the results from the benchmarking were used to extend the classi-
Acknowledgement framework where the framework was not descriptive enough. Rationale behind the benchmark was to identify best practices in annotating so that those could be considered while planning and implementing the annotation framework. Platforms for four types of media were examined: text, image, video, and audio content. These four types are the most commonly found types of media and are expected to be found on a collaborative learning platform.

For each type of media, annotation platforms and applications were examined and conclusions for a collaborative learning platform were drawn. For textual media, highlights turned out to be the most important tool, as means for highlighting were found at every examined platform. For image annotation, the possibility to create polygonal regions or polylines on the annotated image stood out as the most important type of context. Audio and video annotation are closely linked: With a single exception, all of the examined platforms offered only temporal context for video annotation which makes the two types of media identical from an annotation’s perspective. Annotating time intervals and points in time were the most observed types of context. Content of annotations was mainly of textual nature and a single platform allowed annotations to reference other media or annotations. Creating freeform drawings as annotations was observed as well.

To address the information flood users are faced with in a collaborative annotation setting, two approaches were discussed: Different views on annotations and mechanisms for ordering and filtering annotations. Three different views on annotations for textual media, proposed by Marshall et al. [49], were introduced and adapted to different types of media: Displaying all annotations directly on the medium, display annotations and their context detached from the medium, and aggregate annotations to interesting regions. Those possibilities are not mutually exclusive and can be used in conjunction. For filtering and ordering, two approaches were discussed: By the user’s expertise and by rating. For determining expertise, different approaches using different sources of information were introduced, and for rating, both implicit and explicit ratings were discussed.

All findings from the previous chapters came together in the creation of the annotation framework Annoto. Annoto introduces a separation between media-independent and media-dependent functionality: Media-independent functionality is contained at a single location and is shared between all annotation components. Media-dependent functionality is added by the respective annotation component. The support of different annotation types is enabled by encapsulating context and content creation in independent modules which can be arbitrarily combined.

Using Annoto, three annotation components for PDF, image, and video/audio annotation were implemented. The findings from the
benchmarking were used as inspiration for the different Selections and Annotation Types that were implemented. While not every finding of the benchmarking was implemented in the annotation components, the grunt work is done, and implementation of further types of annotations can easily be undertaken.

FUTURE WORK

There are different directions for future work: Work on improvements for Annoto itself and deploy Annoto in real-world applications.

In the short run, there are several aspects where the current implementation of Annoto can be improved. It is very likely that Selections can be made media-independent by introducing interfaces for the exchange of context information. Parts of the framework could be potentially made more generic, such as the current implementation of filtering which only allows filtering by Selections. Due to time constraints, the current implementation for Selections and filtering was chosen instead of a more generic version.

In the long run, there are other conceivable improvements such as multimedia annotation, i.e., the annotation of a document that consists of different media types, such as a text document with embedded videos. This would require Annoto to setup the whole workspace itself and the removal of the dependency on an existing HTML document with pre-defined elements.

Annoto is to be used in a reimplementation of Backstage, a digital backchannel for large class lectures. It allows students to ask and answer questions of other students and participate in quizzes during a lecture. In the new implementation, a lecture consists of more than traditional lecture slides, namely different types of media, such as a page in a scientific paper, a short video of a topic discussed in the lecture, and - of course - traditional lecture slides. Annoto is required to allow annotation for those types of media.

Another vision for an application of Annoto is a digital annotation workspace (tentatively named Annotation Conqueror): Currently, the problem with digital reading and annotation is being limited to a single page which is not the way reading takes place in the real world. In most cases, different pages are bookmarked in some way, or a table that is referenced from different pages is placed next to the current page. A digital annotation workspace would try to solve this problem by enabling the user to view different parts of the same or other documents simultaneously and create connections between them. Similar approaches were examined by Wu et al. [78] and Pearson et al. [61].

Both applications would enable a multitude of studies on aspects of annotations that are insufficiently or not researched at all. The studies which examined the effect of collaborative annotations on learning

---

1 http://backstage.pms.ifi.lmu.de
limited sharing annotations to a small group of people. It would be interesting to study the effect on learning when annotations are shared between all listeners of a lecture. As already touched on in Chapter 5, filters and ordering of annotations have to be studied further. It should be studied in what way filters and ordering of annotations contribute to reduce cognitive load, and if so, which configuration yields the best results. Regarding the digital annotation workspace, it should be examined if the presented assumption really holds: Does space matter when reading and creating annotations?

The future work discussed here is only a small part of possible applications for Annoto. By creating the concept of and implementing Annoto, this thesis laid a solid foundation for applications and studies in the area of annotations, but there is still much work to do before those come to fruition.


[36] Jiun-Chi Jan, Chih-Ming Chen, and Po-Han Huang. Enhancement of digital reading performance by using a novel web-based...


