

MSc project plan:
Intelligent Chat Agent for Q/A on
Programming using StackOverflow as
Knowledge base

130533 - Knut Lucas Andersen (knut.andersen@hig.no)



Master's Thesis Project Description
Master of Science in Applied Computer Science
5 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2015

Avdeling for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Revision history

Version #	Description of change (why, what where - a few sentences)
0.1	First version: Problem description (Chapter 1)
0.2	Second version: Added sections to Chapters 2, 3 and 4. Added Appendices for the Survey and Interview format used in IMT5251 Advanced Project Course (see p. 21 and p. 23). Updated Section 1.6: Planned contributions. Added content to the following Chapters: 4 and 7.

Abstract

When first starting to learn how to program, there is a lot of information to take in. There are a lot of different programming languages, some which have their own rules on how to build and execute the developed program. There are tons of different algorithms and different ways a problem can be solved, not to forget all the different terminologies and semantics that exists within the field. There is also a great amount of online resources on the Internet, ranging from encyclopaedias (e.g. Wikipedia), tutorial sites (e.g. TutorialsPoint, W3Schools, HackThisSite, etc), to online communities (e.g. StackOverflow, CodeProject, etc).

When looking for information, searching for an answer, or looking for a solution to a problem, it is not always that easy to come up with a good question (what defines a good question?). When learning to program in class, the questions may not come right away. Seeing something explained on the blackboard is quite different from actually understanding and doing it yourself. Looking for answers online can become quite time-consuming, since the relevance of the results returned by the search engine vary in a large degree. The answer to the question asked may not appear until page 10 of the returned results.

To help students learning programming, the goal of this thesis is to create a plug-in for the Learning Management System (LMS) Open Edx. The plug-in to be created is an Artificial Intelligent Chat Agent (aka. a ChatBot), and the students can then ask this Chat Agent questions related to programming. The Chat Agent will use StackOverflow as its Knowledge base, meaning the answers presented to the user will be based on the answers posted by users on StackOverflow. By using this Chat Agent, students can ask questions in the same way they would ask a teacher or a classmate, instead of having to use keywords when using a search engine. The goal of this thesis is to see if the Chat Agent can aid students that are learning programming, but also do research on it (e.g. trying to humanize the Chat Agent (Turing Test)). An example of the experiment would be to have two test groups (A/B testing), where the comparison would be on the grades of the students using the Chat Agent vs. those not using it, to see if the Chat Agent had any effect on the students grade.

Contents

Revision history	iii
Contents	iii
1 Introduction	1
1.1 Topic covered by the project	1
1.2 Keywords	1
1.3 Problem description	1
1.4 Justification, motivation and benefits	2
1.5 Research questions	3
1.6 Planned contributions	3
2 Related work	5
2.1 Chat Agent vs. Search Engine	5
2.2 Chat Agents for Learning and Education	5
2.3 Turing Test	5
2.4 Question-Answering (Q/A): What defines a good question?	5
3 Choice of methods	7
3.1 Survey and Interview	7
3.2 A/B testing as an experimental method	7
3.3 Question-Answering (Q/A) model	7
3.4 Quantitative comparison of the students results	7
4 Milestones, deliverables and resources	9
4.1 Table of Contents: Master thesis	9
4.2 Obtaining the desired knowledge	10
4.3 Produced deliverables	12
4.3.1 Hours needed by me	12
4.3.2 Hours by others	12
5 Feasibility study	13
6 Risk analysis	15
7 Ethical and legal considerations	17
A Acronyms and abbreviations	19
A.1 Acronyms	19
B IMT5251 Advanced Project Course: Survey	21
C IMT5251 Advanced Project Course: Interview questionnaire format	23
Bibliography	25

1 Introduction

1.1 Topic covered by the project

The goal of this thesis is to develop an Artificial Intelligent Chat Agent (aka. ChatBot), which will function as a plug-in in the LMS Open Edx¹. This Chat Agent is targeted at students learning programming, and will therefore be used to answer the students questions related to programming. The answers from the Chat Agent will be based on the content found in the online community StackOverflow². Since StackOverflow is one of the many communities belonging to the StackExchange network³, the Chat Agent can later on be expanded to cover information from other communities.

The hope is that this Chat Agent can aid the students in their learning progress, since they can ask the Chat Agent questions in the same way they would ask their teacher or their classmate. They do not have to think about keywords or read through a lot text, since the Chat Agent will try to answer their questions based on the answers it finds in the posted answers on StackOverflow.

From a research perspective, it would be interesting to see if the Chat Agent will have any effect on the learning outcome and what the students think of the Chat Agent, e.g. is it useful or would they just prefer to continue using search engines and look for the answer(s) themselves. As for the Artificial Intelligence (AI) side, it would be interesting to see what can be done to "humanize" the Chat Agent so that it could pass the Turing test⁴. It would also be interesting to look more deeply into the AI algorithms, to see in what way they can be improved or extended to cover a larger base of linguistics.

1.2 Keywords

Intelligent Agent, Chatbot, Natural language processing, Human-computer interaction, Education, Question-answering

1.3 Problem description

Can we find the answer if we do not know the question? The issue with most search engines today is that they are based on taking each search word (the Term) to create what is called a Dictionary (which contains all the words/terms searched for). It then looks through its content (documents, files, multimedia, etc) and searches for each term, and returns those that contain at least one of the terms, ranking the results according to frequency of the given term. Although a given amount of search results is returned, there is no guarantee that the answer searched for are among the returned results. While programming, if an issue occurs, sometimes you can find the answer by using a few keywords, or simply copy/pasting the error message. But what do you search for when

¹Open Edx: <https://open.edx.org/>

²StackOverflow: <http://stackoverflow.com/>

³You can see all the StackExchange communities here: <http://stackexchange.com/sites>.

⁴A Turing Test is a test where a human is asked to converse with a unknown party, and then later on decide whether the party was a human or a computer ([p. 2][1]).

your question is more abstract? What do you search for when you have a question, but are struggling with phrasing it in a way that a search engine can understand? What do you do when you have a question that a teacher or a classmate could easily answer, but the search engine cannot?

With a Chat Agent, you do not have to think about keywords, phrases or "words best describing the problem". You can just ask the question you want an answer to. You also get anonymity with a Chat Agent. You can ask all sorts of questions, no matter how dumb you feel they are, because the Chat Agent is there to help.

A keypoint to remember is that the Chat Agent is intended to function as a help tool (e.g. FAQ FINDER [2] and Bzz [3]), and not as a replacement for the teacher (e.g. CALMsystem [4]).

1.4 Justification, motivation and benefits

With the advancement of programming and the increase of different languages, libraries and functionality, it can be hard for a teacher to cover all topics. It can also be hard for a student to grasp everything at once and understand everything the code does. Although there is a vast amount of online resources on the Internet, finding what you need can take a lot of time, but when using a Chat Agent, you can just ask the question and get the answer right away. Since the Chat Agent will be based on Hidden Markov Model (HMM), it can remember everything previously discussed in the conversation, increasing the chance of it helping the student finding the desired answer. The Chat Agent can also help the teachers, e.g. when running exercises in class, students can use the Chat Agent to find answer to the simpler questions, and then ask the teacher for help on the more advanced and problematic issues.

One example is the paper by [3] where they used an existing ChatBot called Bzz for answering adolescents' questions related to sex, drugs and alcohol. Their study showed that the users used the ChatBot a lot⁵, and that the users felt it was faster and better than information lines and search engines. [5] says that a ChatBot can be easier to converse with due to its anonymity. Furthermore, teachers can look at the conversation logs to see what the students have discussed, to be able to map the problems and see how students learn.

[6, p. 268] did two case studies where they compared the use of ChatBots and e-learning in relation to Information Security. They measured the ChatBot experience qualitatively, and 70% of the users found the ChatBot useful and would use one in the future. However, quantitatively they found no significant difference between those using a ChatBot and those using e-learning. CSIEC (documented in [7]) is a ChatBot developed for learning English, and was tested in [8]. The students achieved a very high score at the exam, but as the authors note, CSIEC was tested only between two tests (and there is also a chance of bias, since one of the authors is also the developer of CSIEC).

⁵"42,217 conversations with the chatbot; thus, an average of 11.3 conversations with each lasting 3 minutes and 57 seconds" [p. 516][3].

1.5 Research questions

- How was the Chat Agent perceived by the user (e.g. using the Chat Agent vs. using a search engine)?
- Did the Chat Agent help the user understand/learn more about programming?
- By using A/B testing, is there a (statistical) improvement from the students using the Chat Agent, vs. those not using it?
- In a given amount of executed queries, how many correct results were presented to the user?
- In what way can the conversation pattern (and algorithms used) be improved to pass a Turing test?
- In what way can the technology (i.e. the plug-in) be improved?
- When retrieving question-answer(s) from StackOverflow, some questions may be closed due to it being a duplicate. Can this data be used in any way to see what is defined by the StackOverflow community as a "good" question?

1.6 Planned contributions

To create an Artificial Intelligent Chat Agent for Question-Answering (Q/A) on programming by using StackOverflow as its knowledge base. This chat agent will function as a plug-in in the LMS system Open Edx. The external resources are content accessible on StackOverflow, libraries for content retrieval from StackOverflow, lexical word mappings (e.g. WordNet⁶), word filtering (to be decided) and available resources from Open Edx.

The Master thesis will be an extension of a prototype currently being developed in IMT5251 Advanced Project Course. This prototype just takes the users question and looks for matches on StackOverflow. The answer that is marked as correct (or the most top-voted one, if no answers are marked) will be returned and displayed to the user. Part of the contribution is to make this prototype more intelligent, by using AI and do research on the implemented technology (e.g. trying to "humanize" the Chat Agent (Turing Test⁷)). Part of the reason for wanting to humanize the Chat Agent is because the target group are students who are learning programming. It would be interesting to see if having this tool available can help them learn and understand more about programming. An attempt to confirm this will be by having an experiment with two test groups (A/B testing), where the comparison would be on the grades of the students using the Chat Agent vs. those not using it.

Since this is such a narrow and specific field, the Chat Agent will be based on the Artificial Intelligence (AI) algorithms Hidden Markov Model (HMM) and Bayesian network (Bayes Net). These are chosen because they have been used for a very long time and there is a great deal of research out there on using these for linguistics and Chat Agents. It could of course be interesting to look at newer or different AI algorithms, but the issue

⁶WordNet: <https://wordnet.princeton.edu/>

⁷For the record, it should be noted that most ChatBots fail the Turing test, and this is extremely hard to achieve. Therefore, being able to pass the test might not be possible, but I want at least the Chat Agent to be humanoid enough so that the users feel comfortable conversing with it.

is that these may not have an adequate amount of research and testing in relation to linguistics. It is therefore in my opinion safer to go with HMM and Bayes Net, to ensure completion of the thesis.

The final part of the planned contribution is to research and analyse the definition of a good question. The StackExchange community has a wide amount of sites, with a high level of professional expertise and strict rules for posting questions. Here, questions can be closed or put on hold, based on the questions asked. Some examples are duplicates, questions that are too broad, or marked off-topic [9], [10]. By analysing the posted questions and those asked by the students, could it be possible that the Chat Agent can help students ask better questions? An example scenario would be a programming lecture teaching incrementation. For a student wanting to learn more about this (presuming the student has no previous knowledge on the topic), a natural question would perhaps be to ask Q1) "What is incrementation?" or Q2) "How to increment?". If you input these questions into StackOverflow, you get approximately 31.600 results for both⁸. Furthermore, the first results may not even be relevant. The natural progression would perhaps be then to add to the question "...in programming". This reduces the returned results to 3.083 for Q1, and 2.715 for Q2. This example can be extended even further, e.g. by adding the programming language (e.g.: C++), the results are now halved, with an amount of 1.429 for Q1 and 1.610 for Q2. This simple example proves that knowing how to phrase a good question can have a real impact of the results you get, and as stated in [11], coming up with a problem can be the hardest part.

⁸This search was executed 29.11.2015, on StackOverflow.com.

2 Related work

- 2.1 Chat Agent vs. Search Engine**
- 2.2 Chat Agents for Learning and Education**
- 2.3 Turing Test**
- 2.4 Question-Answering (Q/A): What defines a good question?**

3 Choice of methods

3.1 Survey and Interview

3.2 A/B testing as an experimental method

3.3 Question-Answering (Q/A) model

3.4 Quantitative comparison of the students results

4 Milestones, deliverables and resources

4.1 Table of Contents: Master thesis

Front page

Abstract

Acknowledgements

Table of contents

Glossary

Acronyms

1. Introduction

1.1. Keywords

1.2. Topic covered/Research area¹

1.3. Problem description

1.4. Research questions

1.5. Methodology to be used

1.6. Justification, Motivation and Benefits

1.7. Limitations

1.8. Thesis contribution

1.9. Thesis structure

2. State of the art

2.1. Chat Agents vs. Search Engines

2.2. Artificial Intelligence (AI) for Chat Agents

2.3. Chat Agents for Learning and Education

2.4. Question-Answering (Q/A): What defines a good question?

2.5. Turing test: Humanizing the AI

3. Methodology

3.1. Hidden Markov Model (HMM)

3.2. Bayesian network (Bayes Net)

3.3. A/B Testing

3.4. Survey and Interview

¹Although it is called "Topic covered" in this report, it may be more appropriate to call it "Research area" in the Master thesis.

- 3.5. Research Design
 - 4. A/B Testing, Surveys and Results
 - 4.1. A/B Testing
 - 4.2. Interaction with the Chat Agent
 - 4.3. Statistical comparison of the students results
 - 5. Discussions
 - 5.1. Data and Testing
 - 5.2. Artificial Intelligence (AI) Methods
 - 5.3. Implementation Architecture
 - 5.4. Chat Agent vs. Search Engines
 - 6. Conclusion/Summary²
 - 6.1. Overview of main results
 - 6.2. Further work
- Bibliography
- A. Data sets/Statistical Overview
 - B. User Survey
 - C. Interview Questionnaire Format
 - D. Application Screenshots (Interaction with the Chat Agent)
 - E. Miscellaneous information

4.2 Obtaining the desired knowledge

The most important key element in this thesis will be the development of the Chat Agent, since it is the focus of my thesis. Preliminary work has already been done in the course IMT5251 Advanced Project Work, where the prototype is under development. The prototype runs in Open Edx, as an XBlock. Put shortly, the XBlock runs as an Fragment in Open Edx, allowing developers to add their own content which then can be re-used in multiple systems³. As previously mentioned, the prototype developed is a simplistic version, meaning there is no AI. The prototype will simply take the users input (the question) and copy/paste it to search for matching questions on StackOverflow. It will then return the first result, where the answer the user is presented with will either be the answer marked as correct, or the top-voted answer (if no answer is selected as correct by original poster). Students from the 1. and 2. year (who are learning to program) will test the prototype in the beginning of December, and afterwards fill out a survey about this experience (this survey can be seen in Appendix B). In addition, some will test it out in the data lab while I observe them, and afterwards they will take part in an interview

²Whichever is the appropriate format for the Applied Computer Science Master thesis.

³For more on XBlock, see <http://edx.readthedocs.org/projects/xblock-tutorial/en/latest/overview/index.html>

(the interview format can be see Appendix C).

This means that most of the development needed during the master thesis will be to implement the AI and for it to be able to use WordNET for semantics. This would probably take a whole calendar month, but it depends on the actual hours invested. If I work between 6-8 hours each day (40+ hours each week), then this should be at least operational at latest mid-February. The reason for this extended time is to ensure I have time to setup and test the AI properly before allowing students to test them to ensure the collected data is valid. When it comes to equipment, I already have most of what I need, since I am already working on the prototype (development is done in Arch Linux). I also have a USB stick with Arch Linux installed, so that I can work on my laptop in case something should happen to my Desktop. I have also acquired a student license for PyCharm Professional⁴ which is valid for 1 year (until 25. November 2016). I am also aware of people that have the required knowledge who I can ask for help, such as my supervisor Simon McCallum, Sule Yildirim-Yayilgan, Mariusz Nowostawski and Rune Hjelsovold.

The user testing will of course require students to be willing to partake and test the Chat Agent. The plan is to use A/B testing, so that only part of the students have access to and can use this Chat Agent. One thing that may affect the end results are the knowledge level of the users, which means that even if there is an increase in the end results, it may not be because of the Chat Agent. The solution to account for this issue is by having the users also grade their own knowledge level, from being novice to expert (e.g. having programmed for years). Although there is not a set time limit for required use, if the Chat Agent is not in use, there will not be enough test data. However, the Chat Agent will be available from around February to May, and participation may not be that high if there is a weekly requirement for usage. A solution could be to require the participants to at least use the Chat Agent for at least 10-20 hours each calendar month. If there then are 20-30 participants, that means the Chat Agent will have a usage of 200-600 hours each month. Which in turn should provide a good amount of test data. There will be at least three surveys for the participants, the first when they start using the Chat Agent, the second when testing is halfway and the last at the end, to see if there is a correlation between usage, their results and the students own observations.

⁴ PyCharm: <https://www.jetbrains.com/pycharm/>.

4.3 Produced deliverables

4.3.1 Hours needed by me

Product	Time (calendar)	Time ('man-hours') ⁵	Version #	Notes
MSc thesis report	January - May	125-200 Hours	v0.1	Draft should be presented to supervisor monthly
1) Chat Agent (AI research)	January - February	50-75 Hours	v0.1	Should be available at latest Mid-February for course start-up
2) Chat Agent (AI development)	January - February	50-75 Hours	v0.1	Should be available at latest Mid-February for course start-up
Meeting w/ supervisor	January - May	10-20 Hours	v0.1	Once a week, estimated between 30-60 minutes
Analyse questions on StackExchange	February - May	50-100 Hours	v0.1	One way of doing this would be to use the Chat Agent to e.g. retrieve questions that are closed
Read various scientific papers	January - May	50-100 Hours	v0.1	Read up on the latest published papers to keep up-to-date
Process student surveys	Jan/Feb, March/April and April/May	10-30 Hours	v0.1	Surveys delivered by students participating in the A/B Testing
Process Student/ Chat Agent interactions	February - May	60-80 Hours	v0.1	Should at least try to keep a steady update on the current data on a weekly basis

4.3.2 Hours by others

Who?	Time (calendar)	Time ('man-hours') ⁶	Version #	Notes
Supervisor	January-May	10-20 Hours	v0.1	Once a week, estimated between 30-60 minutes
Students	February-April/May	40-80 Hours	v0.1	Students interaction with the Chat Agent (time per person)
Students	Jan/Feb, March/April and April/May	1.5 - 3 hours	v0.1	Filling out survey/questionnaire (per student)

⁵Total time spent during set period.

⁶Total time spent during set period.

5 Feasibility study

Why can this project be completed in time?

Resource feasibility

- Technical resources
- Skills and competencies relevant in solving the problem
- Resources for generating data to be analyzed

Schedule feasibility

- Size and complexity of tasks, deliverables, and milestones
- Headroom for the unexpected

E.g. project in Adv. proj. work, comparison to other projects, attempt to answer research question, etc.

How to solve the issues/problems in the project

6 Risk analysis

What can possibly go wrong when you do your project?

How do you intend to reduce impact of/solve these problems?

Objective: Identify and assess factors that may jeopardize the success of your project

- Probability
- Consequence

Often caused by uncertainties:

- Lack of knowledge
- Complexity of the project
- Inherent randomness

Consider the specific risks - not the general ones like getting sick, insufficient feedback from supervisor etc.

7 Ethical and legal considerations

There are no legal considerations in this project. All user data be anonymized, and the only data logged by the Chat Agent is the questions that the students ask, and the answers they are presented with. The greatest ethical concern in this thesis is the invested amount of hours testing the Chat Agent. The students will have projects, deadlines and exams to relate to, so requirement of 10-20 hours of usage each month may be too much. On the other hand, if the Chat Agent is not used at all, there may not be enough test data to prove whether or not it had any effect on the students learning curve. If one were to set a timed requirement, it could increase the chance of stress and performance anxiety, ending with students declining to participate or withdrawing before the testing is completed. It is important for this thesis both that the Chat Agent is used, but also that the users want to use it! Another concern in relation to time is that there is no guarantee that the Chat Agent have any beneficial effects on learning. The results may show that it is faster or easier to use search engines, meaning students have lost time they could have invested elsewhere.

For the validity of the end results, there is also the question of whether or not the Chat Agent actually had any effect. Students past knowledge can effect the average outcome, for better or worse. During the thesis, the testing will be conducted by using A/B testing, meaning only a selected group will have access to the Chat Agent. If a large part of the group have a lot of knowledge from before, the end results may be mostly False-Positives. Because even though the data show improvement, they improved because they build on previous knowledge, and not by use of the Chat Agent. To catch this, one of the questions for each test user will be to grade their own knowledge level to scale the end results. There is also the issue that students who might benefit from the Chat Agent, will not be able to use it. Novice students, with little to none programming knowledge may find it unfair that they are not the primary selected group. However, by not selecting by type (e.g. knowledge) one can avoid affecting the data and the end results.

A Acronyms and abbreviations

A.1 Acronyms

AI Artificial Intelligence. 1, 3, 10–12

Bayes Net Bayesian network. 3, 4

HMM Hidden Markov Model. 2–4

LMS Learning Management System. i, 1

Q/A Question-Answering. 3

B IMT5251 Advanced Project Course: Survey

C IMT5251 Advanced Project Course: Interview questionnaire format

Bibliography

- [1] Russell, S. J. & Norvig, P. 2013. *Artificial Intelligence: A Modern Approach*. Pearson Education, 3 edition.
- [2] Burke, R. D., Hammond, K. J., Kulyukin, V., Lytinen, S. L., Tomuro, N., & Schoenberg, S. 1997. Question answering from frequently asked question files: Experiences with the faq finder system. *AI magazine*, 18(2), 57.
- [3] Crutzen, R., Peters, G.-J. Y., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. 2011. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *Journal of Adolescent Health*, 48(5), 514–519.
- [4] Kerly, A., Ellis, R., & Bull, S. 2008. Calmsystem: a conversational agent for learner modelling. *Knowledge-Based Systems*, 21(3), 238–246.
- [5] Knill, O., Carlsson, J., Chi, A., & Lezama, M. 2004. An artificial intelligence experiment in college math education. *Preprint available at <http://www.math.harvard.edu/~knill/preprints/sofia.pdf>*. Last Accessed: 25.09.2015.
- [6] Kowalski, S., Pavlovska, K., & Goldstein, M. 2013. Two case studies in using chatbots for security training. In *Information Assurance and Security Education and Training*, Dodge, Ronald C., J. & Fitcher, L., eds, volume 406 of *IFIP Advances in Information and Communication Technology*, 265–272. Springer Berlin Heidelberg.
- [7] Jia, J. 2009. Csiec: A computer assisted english learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems*, 22(4), 249–255.
- [8] Jia, J. & Ruan, M. 2008. Use chatbot csiec to facilitate the individual learning in english instruction: A case study. In *Intelligent Tutoring Systems*, Woolf, B., Aïmeur, E., Nkambou, R., & Lajoie, S., eds, volume 5091 of *Lecture Notes in Computer Science*, 706–708. Springer Berlin Heidelberg.
- [9] Stackoverflow.com. 2015. What does it mean if a question is "closed" or "on hold"? - help center. <http://stackoverflow.com/help/closed-questions>.
- [10] CommunityWiki. 2015. What is a "closed" or "on hold" question? <http://meta.stackexchange.com/questions/10582/what-is-a-closed-or-on-hold-question>.
- [11] Lucky, R. 2011. When the problem is the problem. <http://spectrum.ieee.org/at-work/innovation/when-the-problem-is-the-problem>.