NTNU – Trondheim
Norwegian University of
Science and Technology

Guiding Learning by Creating an Adaptive User-Model of Knowledge, Memory and its Decay

Anders Oliver Dahlin

Master of Science in Informatics
Submission date: June 2015
Supervisor: Asbjørn Thomassen, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

Background As time pass, education has been an ever growing enterprise. The past 150 years obligatory education has grown from almost non-existent to a project that takes most of our childhood. Adding more years to an already long education is out of the questions, but there are other improvements that can be done to make sure our children learn what they need to know.

Objective The main objective of this paper is to find out how computer systems can help this improvement process. By adapting the learning system based on each students needs and prior knowledge, can we guide the learning process so that they only need to work on the topics where they have the necessary dependencies memorized already? There is also an attempt to find weather we can use existing cognitive architectures(mainly ACT-R) to achieve this.

Research method First, we discuss different educational paradigms and research done on what the best learning experience is like. Second we go through different theories surround how memory and forgetting works. Finally we go through what a cognitive architecture is and how ACT-R works, to then translate that into a the issue at hand.

Results It was found that there are big gaps between the theories surrounding how people best to learn new things and how it is done in schools. When people learn something new it is put inside their working memory and over time rehearsed so it gets consolidated in long-term memory as well. The

more items in working memory at the time the more competition there is for rehearsal, that's why you learn better when focusing on one thing at the time.

Even after a memory is consolidated in long term memory it is not necessarily permanent. The brain deletes useless memories all the time. It is not known exactly how this happens, but it is usually based on a function of time. There is also evidence of similar memories conflicting with each other. Forgetting things is not a weakness, it's just how humans work. It is therefore important that in an attempt to help people learn we should not believe memories last forever.

Taking this into account it was found that an adaptive system for helping people learn need to try and predict when users will forget things and then trigger a recall in order to consolidate the memory even further. It was also found that it must use the model of what the user knows and don't know to make sure the user is allowed to only work on topics they feel comfortable working on. Not too hard, not too easy.

An argument was raised that we should not attempt to force people to remember everything forever as human attention is a commodity and as technology advances we already have different tools available for accessing long term storage of information. Only the information useful to the individual needs to be remembered after it has been processed.

ACT-R works by using time-stamps of when memories have been accessed and creating an activation value based on those. That activation is then combined with a signal from it's neighbours to create a probability for the mind being able to recall that information again at any specific time.

By creating a tree or graph of different declarative information nodes and their dependencies we can use the activation function from ACT-R to predict users ability to remember different topics and use that prediction to guide their learning.

Sammendrag(Norsk)

Bakgrunn Utdanning er et foretak som har vokst mye gjennom tidene. De siste 150 årene har obligatorsk utdanning vokst fra nesten ikke-eksisterende til et prosjekt som tar det meste av barndommen vår. Å legge til flere år på en alleredet lang utdanning er ikke lurt, men det finnes andre forbedringer som kan bli gjort for å passe på at barna våre får utdanningen de fortjener.

Mål Hovedmålet til denne oppgaven er å finne ut om et datasystem kan hjelpe til med denne forbedringsprosessen. Er det mulig å lage et adaptivt læringssystem som tar hensyn til brukerens behov og tidligere kunnskaper for å guide læringsprosessen mot temaer der de har de nødvendige forkunnskapene? Vi forsøker også å se på mulighetene for å bruke kognitive arkitekturer(hovedsaklig ACT-R) for å oppnå dette.

Forskningsmetode Vi skal diskutere forskjellige utdanningsparadigmer og forskning gjort på hva som gjør den beste læringsprosessen. Så skal vi gå gjennom forskjellige teorier rundt hukommelse og glemsel fungerer. Til slutt går we gjennom hva kognitive arkitekturer er og hvordan ACT-R fungerer for å så overføre det til problemstillingen vår.

Resultater Vi fant at det er store avstander mellom teoriene rundt hvordan mennesker best lærer nye ting og hva som blir gjort i skolen. Når mennesker lærer noe nytt, blir det tatt opp i korttidshukommelsen og over tid øvd inn slik at det forplanter seg i langtidshukommelsen også. Jo flere forskjellige tanker du har i korttidshukommelsen om gangen jo mer konkurranse har tankene dine om øvingskapasitet. Du tar derfor til deg kunnskap bedre når

du fokuserer på kunn en ting om gangen.

Selv etter et minne er blitt forplantet i langtidshukommelsen betyr ikke dette at det vil vare for alltid. Hjærnen kvitter seg med minner hele tiden. Det er ikke visst nøyaktig hvordan dette skjer, men det er vanligvis en prosess basert på tid. Det er også bevis som sier at like minner kan overlappe med hverandre. Å glemme ting er derfor ikke en svakhet, det er bare slik vi fungerer. Når vi så prøver å lære nye ting til andre er det viktig at vi tar hensyn til dette. Ikke forvent at elevene våre vil huske ting for alltid.

Når vi tok hensyn til dette, fant vi at et adaptivt system for å hjelpe folk å lære trenger å prøve å forutse når brukeren kommer til å glemme noe. Da kan brukeren få hjelp til å forfriske hukommelsen noe som vil hjelpe til å styrke forplantningen til det temaet i langtidshukommelsen. Vi fant også at med en modell av hva brukeren vet og ikke vet så kan vi passe på at den bare får jobbe på temaer den føler seg komfortabel å jobbe med. Ikke for lett, og ikke for vanskelig.

Det kom også et argument rundt at vi ikke burde prøve å tvinge folk til å huske alt mulig for alltid. Menneskelig oppmerksomhet er en mangelvare, og slik teknologien går forover har vi mange andre verktøy for å ta vare på informasjon over lang tid. Bare informasjonen som er viktig for individet trenger å bli husket etter at det har blitt prosessert.

ACT-R fungerer ved å lagre tidspunkter for når et vist minne har blitt aksessert. En aktivasjonsverdi blir konstruert ut ifra denne dataen og så kombinert med signaler fra minnet sine forbindelser med andre minner. Dette blir brukt for å lage en sannsynlighet for om den spesifike brukeren klarer å huske informasjonen på et spesifikt tidspunkt.

Ved å lage en graph av forskjellig deklarative informasjonsnoder og deres avhengigheter kan vi bruke aktivasjonsformlene fra ACT-R for å forutse når brukerens evne til å huske forskjellige temaer ikke holder mål. Denne informasjonen kan så brukes for å guide læringsprosessen.

Preface

I remember as a child, I used to love school. I was interested in learning new things, and that was reflected in my results as well. In the last years of elementary school I ran into a problem though. The curriculum was mostly repetition of what we had learned earlier. Repetition is not a bad thing in itself, but I knew most of what we went through in class already and felt bored. I could have spent this time learning new things, but that's not how school worked back then. When I joined middle-school I was met with the opposite problem. The classes went through too many new things for me to be able to keep up, especially when I was getting used to not having to pay attention. I didn't get bad grades, but there was always something wrong that I hadn't learned well enough. After we finished a topic we would have a test, but if we didn't know everything well enough on that test we still began working on a new topic the next week. Small gaps in knowledge become bigger over time, and eventually you get a new topic that builds on what you should have learned last year, but didn't quite get right back then. And the work you have to go through doubles.

The lesson here is that customizing what you need to work on after what level you are at, is much more logical than thinking every student in a class will be at the same level throughout their education. We want our kids to learn what they need to know to be competent adults, but that is not what our schools focus on. They just want students to cram knowledge into their heads before a test so it will seem like they get good results, when most of what they have learned is forgot a month or a year later.

I also remember a feeling of being lost in a huge list of topics and subtopics that is put together in seemingly arbitrary ways throughout 15 years of education. I could never compare what I knew with what I was supposed to know. There has always been a distinct lack of scale, it was impossible for me as a child to imagine how much more I was supposed to learn when I became 12, 15 or 20.

I believe some way for students to compare themselves with where they are supposed to be at the end of a school year. To compare how far they has come in relation to where they will need to be when they begin in high-school. To see how much progress they are making. To see if they are ahead or behind the rest of their class. These kinds of tools would help any student keep motivated to work, and help them focus their attention where it is most needed.

A seemingly endless stream of mindless work without context is not a good way to motivate someone to do well in school.

Contents

Chapter 1

Introduction

All knowledge is built on other knowledge like bricks in a brick wall. When you try to gain knowledge of one of the bricks, it is easier to fit it in if you have bricks underneath to support it. Sometimes however those bricks are missing, and it's hard to know what bricks you need and where you can get them. The human brain can't remember everything, so finding these bricks can be a real hurdle to your learning experience. A computer however can remember everything forever, but it doesn't currently know what you know. Is there a way to tell it? Can it foresee how well you have learned each brick of knowledge and when the natural decay of memories will occur? Can we model the different bricks of knowledge in a way that lets us see all the relations between them, and from that relation-graph find out what knowledge is needed to better understand the actual knowledge you want to learn?

I imagine a system where users only have to log in and pick a topic they want to master. The system goes through the tree of knowledge and finds the topic best suited for them. Learning materials for this topic is provided, and at the end they do a test to see if they have understood. Every time they master a topic, new topics are opened for learning and eventually they reach the target chosen. Only the topics whose prerequisites have been mastered are open for learning and that's how we make sure that they are not forced to work on something they don't have the know-how to do. It should also make sure they don't have to work on something that they know already.

The system will recursively go through the pieces of knowledge that it thinks you need to know first to learn the topic chosen. Check each section for the users prior knowledge, and guess the likelihood of the users having forgotten it based on how well they knew it and how much time that has past. Then create a systematic plan of attack for working up to the knowledge targeted. All in while providing tips and information from other users and/or teachers working on the same topics.

Chapter 2

Background and Motivation

I've been interested in different computer based learning systems for a long time. I've used many of them myself and heard talks and presentations about others. Khan academy especially has done an amazing job. However from all I've seen there always room for improvements. That's why I wanted to do some looking into what could be done to fill that space.

There is no doubt that computer based learning systems is the way of the future, you have simulations to train soldiers and plane pilots, you have software to help learn languages. There are a bunch of tools and lectures on-line to help with maths or physics or any other hard science subject we learn at school.

One of the biggest hurdles is understanding how the human mind work and how we learn or forget, But we are starting to get ways of modelling these things that has achieved good results in several different simulated tasks.

Gamification is also related to making a good learning platform. Games have an unique ability to engage people, and that is exactly what we want to do when teaching. By making each task manageable and rewarding the student for each success and providing a competitive environment where everyone can feel successful, it is believed we can make learning both enjoyable and accessible for all.

Chapter 3

Research Questions

- How can we create an adaptive learning system?
- Can we use existing cognitive architectures like ACT-R to make the system imagined work?

3.1 Adaptive learning

The main goal is to make the process of gaining new or refreshing old knowledge as painless as possible.

Learning something new comes with challenges. Sometimes the idea is too abstract to grasp without concrete examples, sometimes you don't know what is happening because you lack understanding of underlying principles. Sometimes you are swamped by the sea of topics that surround what you are trying to learn, unable to focus on any one thing. Sometimes you just don't like the way the text-book you are given explains it.

What can be done to make a learning process adapt to individual preferences and backgrounds? What can be done to make it give the learner motivation to continue? What can be done to make a daunting task seem more manageable? What can be done to make sure we don't have to go through the overhead of wading through useless information before you get to the topic we want to focus on?

These are some of the questions I ask myself when I think of adaptive learning, and I am going to attempt answering some of them.

3.2 Cognitive architectures

"ACT-R is a cognitive architecture: a theory for simulating and understanding human cognition. Researchers working on ACT-R strive to understand how people organize knowledge and produce intelligent behavior. As the research continues, ACT-R evolves ever closer into a system which can perform the full range of human cognitive tasks: capturing in great detail the way we perceive, think about, and act on the world."

From the ACT-R web-page on the Carnegie Mellon University website.

ACT-R has had a lot of success with simulating human cognition. I will attempt to use it as an inspiration for my own system.

Chapter 4

History

4.1 Teaching

You could argue that our ability to reason and think logically, is why we are as advanced a species as we are. However if you put a child born today in a family living 2000 years ago, that child will not be any more advanced than the other children at that age. It is the teaching of current knowledge to new generations that is the main reason for why we are as advanced as we are today. Each generation does not achieve much on their own, but they have the knowledge of their ancestors to build new information on. This allow us to achieve progress as a species.

Getting better at teaching our new generations is therefore one of the most important ventures we can undertake, and there is a lot of room for improvement. Schools as we know them is a recent invention, but institutions for learning have existed a long time. In ancient times only people from certain backgrounds where allowed or had the money or free time to become scholars. Only a small percentage of the population where literate, they were mostly taught in logic, religion, basic science and some sort of physical art. During the middle ages the monasteries were the centres of learning in Europe. The monks taught Latin and how to write, but also sciences. In the 18th century, literacy grew substantially. Each country had their own circumstances leading to their own ideas of how to teach, who to teach, and what to teach them, but by the 19th century free education for all started to grow. It became compulsory in England for children up to 10 in 1880, and compulsory up to the age of 16 in 1972.

We have come a long way in improving how our children are taught, but adding more years to an already long education might not be the best way to improve it further. As it stands, each year at school has a curriculum that students have to learn. The goal of the school is to make sure as many students as possible know as much of the curriculum as they can when taking the exams at the end of the year. However students ability to do exams do not decide what they will know when their education is finished. As someone nearing the end of my schooling I asked some of my classmates what they felt they had gained from the 18 years we have spent in school. They talked about how they learned how to socialise as a human being, they said they got an overview on how the world works and an idea on how the past was like compared to the present. We learned how to gain new knowledge and critical thinking. No one claimed to know what the curriculum was in the 10th grade, or that they would be able to pass the tests they had back then now. The only specific knowledge they had retained were the things they had continued to use after they learned it. Arithmetic, English language, statistics, probability and some algebra. It's important to note that these people all went to study computer science. They remembered the things they had use for in their later studies.

Long term memory is not forever, students learn things once, and then forget it eventually if it was never useful again. If we accept that this is the reality, we can't expect students to remember everything from the curriculum two or more years earlier. When something new based on old topics is brought up, students need help finding out how to catch up to where the new curriculum expect them to be, so that everyone is on an equal sturdy footing when learning about new topics. But after students learn something, how long do we expect them to remember it? why? We need to ask ourselves to what extent do students have to learn the curriculum. Gaining knowledge about the world helps you put your view on things in perspective, but how long do you need to remember it before that benefit stops being relevant? Is it good enough to remember it for a day on the exam, a month? A year? What is the difference if the information itself had no practical use during the time you remembered it?

Chapter 5

How to achieve an adaptive learning process

5.1 Related work

5.1.1 Flashcards and spaced repetition

There are many programs that implement flashcards and spaced repetition to help people learn. The idea is to allow a user have a set of items she wants to learn. This can be glossary in a language, pronunciation of Chinese characters, Doctor Who trivia or any other information that can be represented as a question and an answer. An example is a set of words in Spanish. The flashcard says: "gato", if you remember that this means cat, you have succeeded and you get a new question.

The traditional system is physical. If you remember the answer to the question presented, you are responsible for judging your own performance. and place the card in the appropriate level according to your success or failure. This lets an individual schedule her learning without much hassle.

When spaced repetition software(SRS) started becoming popular the need to keep things simple for the user was still important, but the computer could take on complex tasks in the background.

One version of such a system is Anki. When you get a question from a card in Anki, you rate the difficulty you had answering. If you think it is easy the algorithm will schedule that item further ahead than if you thought it was difficult to remember. This way, instead of having set levels with set intervals for practice. Each item gets a customized time interval depending on how many times you have succeeded in a row and how difficult you rated the word upon answering it.

Another system called Duolingo takes it one step further. Questions in Duolingo can also be audio files. And more complex questions can have several correct answers, when for example the grammar is flexible. It also uses pictures in combination with words to help the user remember.

A third system of note is Memrise. In Memrise users create their own courses freely. They can combine audio, video, text, pictures and even hints to the different questions that are practised. But the answers are always text. This allow for courses about anything to come into existence as long as there is an interest in it. On their website you can find courses to learn anything from Dothraki or dragon tongue, to chess moves or pokemon. The negative side of user created content is that you get a lot of duplicates. Each version has their own strengths and weaknesses, and because of the lack of standardization there is little chance for any of the faults to be fixed after the course has already been created.

Each course is divided into levels. Here levels are not how well you know something, but rather what order the user wants you to learn them. Each level has a number of items, it's up to the creator how to group them. But a level can also contain multimedia for general learning purposes, entertainment or lessons that is not easily explained as a question-answer. Memrise uses the analogy of each item you learn being a seed, that needs to be planted and then watered regularly. When watering your seeds you have to answer the question given correctly, but at first it will give you a multiple choice version of the question, to ease you into it. Later you have to be able to write out the answer yourself. When planting you are given all the information about the item the creator has put into the system. The native word, the translation, alternate spellings, the pronunciation, how it's written. Anything that's worth knowing. However there is also a second feature to help you remember. Any user may add a memorization rule to any item in any course. If you have a really good way to remember a word or a chess move, you can write it down with a picture if you want to and have it listed for every other user to see. If someone else find your way of thinking helpful, they give it a thumbs up and every rule get sorted according to that rating. This takes some of the responsibility off the creator and let others contribute with their knowledge or experience.

5.1.2 Other on-line courses

Another on-line learning course is called Codecademy. This tool is created for people interested in learning to code programming languages like HTML, JavaScript and Python.

The course is set up so that you start at the very basics. Each section of one lesson is small and easy to wrap your head around, and there are only one or two new things to think about at the time. The new topics are demonstrated for the users, and then the they have to replicate what was demonstrated in a small challenge. When the user succeeds with their task, they start working on the next section that builds on what they just learned. This way you are able to focus on the new things you are learning as you develop your skills. And suddenly you are an expert at programming without even noticing.

Spaced repetition is not used here so remembering what you learn can be a problem if something you learned 30 minutes ago have not been used since. However the teaching itself utilizes the strength of a computer to its fullest by modelling a real scenario and letting the users see the effects of their input, even if it's wrong.

Khan academy is one of the most advanced learning systems readily available today. It has helped show the world how the right presentation can help students learn anything by themselves. And how gamification of the learning process is more than just a buzzword. In 1984 Benjamin Bloom researched the difference in students that are pulled out of class and given individual tutoring instead.[9] The students interest and aptitude was of similar strength and the amount of time for instruction was the same for both groups, but in the end those with private tutoring achieved in average better score than 98 percent of the control group. One on one tutoring is effective, but expensive. Khan academy tries to bring the advantages of one-on-one tutoring into an affordable package.

The idea is simple. When met with a new topic, a student is shown a video.

In this video Khan draws diagrams and examples of how this topic works, while a voice-over explains what's going on. The student can fast forward, jump back or pause the video at any time to get a repeat of something she didn't quite get the first time. This give her a tool to spend as much time as she wants figuring out the problem, without having to ask a teacher to repeat himself. When finished, there are practice problems that are generated for each topic. These are never the same, so the student need to learn how to solve the problem, not just remember the answer.

When practising a topic in Khan Academy you get 5 different questions you need to answer correctly. Doing so will let you achieve a higher level at that topic. However you can do something called a mastery challenge to get a question from five random topics you have practised. If you answer correctly you will achieve the level of mastered. It seems to sometimes add already mastered topics in the mastery challenges, this make sure that if you forget how to do that topic over time, you will lose the status of mastered.

One success story, tells about a teacher Kami Thordarson who turned the traditional system upside down.[7] [8] She replaced some of her own lectures with Khan videos, whom the students can watch at home. Then, in class they sit together and solve the practice problems. This way it is the lectures that are viewed on the student's own time, while the homework is done at school. This makes sense as you are more likely to need help while actually working on the problems. Khan Academy also provides a dashboard for the teacher to keep track of all his students. He can see if someone is stuck, he can see who has watched the videos they where supposed to watch. This way he can give help where it is needed, and guide the classroom discussion to fit what the students are working on.

The success is not attributed to high quality videos or perfect website design. The videos are well made, but they could be better. There could be more tools than just the videos for learning too. There are still work to be done. But Khan academy is a proof of concept. This is worth developing.

In Norway, Gyldendal[4] is also developing such a system for use in schools.

They are currently building a knowledge graph of mathematics in order to map what topics build on each other. Their goal is to let the students work on the topics that are on their own level, not too difficult and not too easy. The old problem of a teacher only teaching for the students in the middle range of the class has been a tricky one. The students on the lower end gets left behind and the students ahead of the class get bored to death. With a customized adaptive system for providing students topics to work on, this problem will hopefully be solved.

5.2 Research

5.2.1 Memory

Memory is not well understood, but we have some idea of how it is structured.[16] The distinction between short-term memory and long-term memory has existed since the 19th century. Short-term has a limited size(memory span), you can only keep between 5 and 10 memory chunks active at any one time, and the duration is limited to a few seconds to a minute. Long-term on the other hand don't have any limits to capacity or duration. But rather have limitation on accessibility.

Forgetting(STM) There are two main theories for why we forget short term memories. The first is trace decay and relies on the idea that memories leave a trace in the brain that decays over time. It suggest that you can only remember something between 15 and 30 seconds unless rehearsed.

Everyone agrees that memory tend to get worse over time, but there is no direct proof that it is decay that causes this memory loss and not the events between learning and recall. Because of the Short term memory's limited capacity it is believed that new information "pushes out" old information.[1]

A typical study conducted have participants listen to a list of words and attempt to remember them later. They are allowed to recall the words in any order they want, but the findings show that the first words and the last words in the list are more likely to be remembered than the rest.

This is explained as the primacy, and the recency effect. The first words heard was remembered better because they were rehearsed more frequently. At the time they where heard they had less other words to compete with in the limited capacity of the short term memory, this makes them more likely to be transferred to the long term memory.

The recency effect is caused by no more new words being added to STM, so that the words already in STM at the end of the learning process are rehearsed more than those who were ”pushed out” during the listing of the words.

Forgetting(LTM) There are many theories on why we forget things from the Long-term memory, and they all have merit. One theory is called Interference and is based around different memories interfering with each other. This is more likely to happen when the different memories are similar.[5]

Destin, the man behind the Youtube channel ”Smarter every day” did an experiment with a bicycle that had the steering work backwards. If you turned the steering left, the wheel turned right, and vice versa. He attempted to learn to ride this bicycle, and in the beginning it was impossible. He could not go more than a meters length before having to support himself with his feet. Many people tried to ride it, but no-one could. It took him 8 months of practice to finally be able to ride it, his son learned it in 2 weeks. When he then attempted to ride a normal bicycle, he couldn't do it. He kept falling off until about 20 minutes had passed when he suddenly managed to use a normal bicycle normally again.

This could be explained by his old knowledge of how to ride a bicycle interfering with learning to ride the one that worked backwards. Why it took him 20 minutes to be able to ride a normal bike again after this experiment could also be explained by the two skills interfering with each other, but he didn't have to learn to ride a normal bike again from the beginning. He just had to remember it.

Another theory is lack of consolidation. When we experience something new it takes a certain amount of time for the changes in the nervous system to consolidate so it is remembered properly. This theory has value because of it looking at the process as a physical event. We are still uncertain exactly how the brain alters the neurons to store the memories.[10]

Finally we have Retrieval failure, which states that memories forgotten are still stored in memory, but we are lacking the contextual cues necessary to access them. Whenever we take in new memories we store information about the situation we are in as well. This can be external sensations like smell or being a specific place or it can be internal feelings, mood or medical states. We associate these cues with the memory and that helps us remember.

This is supported both by scientific experiments as well as everyday experiences.[12] You might not remember much from your time in Elementary school or example, but if you went to the school building you would suddenly be flooded with memories from the time you spent there.

This is also relevant to what Destin experienced when he was unable to use a normal bicycle after spending months on a reverse one. The knowledge of using a normal bike was still there he just needed to access the right associations.

5.2.2 Forgetting curve

In 1885, the psychologist Herman Ebbinghaus came up with a concept called the "forgetting curve".[20] He demonstrated that over a short amount of time, there is a dramatic loss of retention in any knowledge you have just finished consuming. Since his time, many other researchers have confirmed that the retention of new information decay rapidly, unless it is reviewed. An example from a study about the retention of CPR skills after training show that after 3 years, only 2.4 percent where able to perform CPR successfully.[21][22] Another recent study of physicians taking a tutorial they rated as very good or excellent showed mean knowledge scores increasing from 50 percent before the tutorial to 76 percent immediately afterwards. However, score gains were only half as great 3–8 days later. After 55 days there were no significant knowledge retention measurable.[15][22]

Different ways of learning something give different results. Having a topic demonstrated would give you more retention percentage than just reading a text, experimenting/trying it yourself would be even better. The problem is that not every topic is as easily practised by a new learner, and no matter what method of learning you use there will still be a significant loss of retention over time.

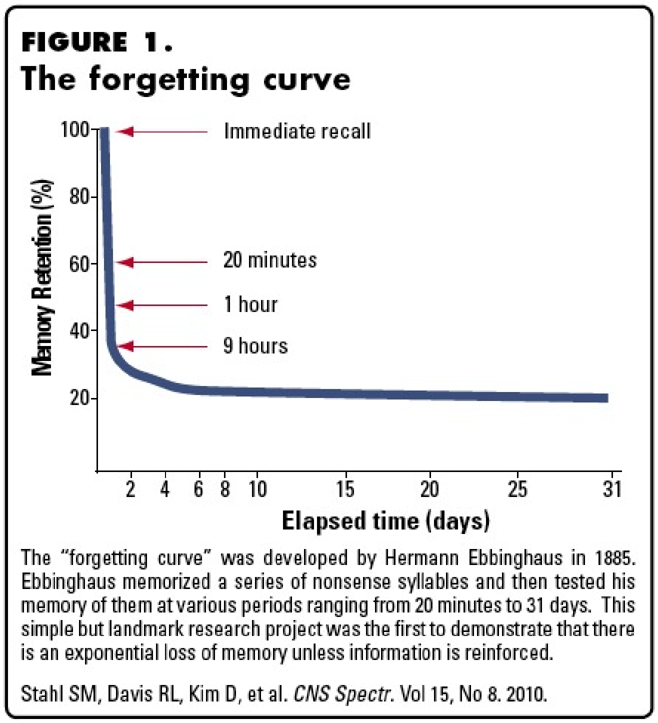

Figure 5.1: Illustration of the forgetting curve. Taken from "Play it Again: The Master Psychopharmacology Program as an Example of Interval Learning in Bite-Sized Portions."

This is where spaced repetition comes in.

5.2.3 Spaced repetition

Spaced repetition is a technique for learning where you repeatedly drill the subject you are working on in gradually growing intervals. There is a phenomenon in psychology called the spacing effect that states that few repetition over a long period of time is more effective than many repetitions over a short period of time.[20]

This technique is mostly used for learning through flashcards, and was quickly implemented as software based solutions as personal computers became popular. A computer can automatically keep track of arbitrary amounts of flashcards at the time, scheduling when each topic needed rehearsal. Making the whole process much easier for learners. There are two popular ways of scheduling.

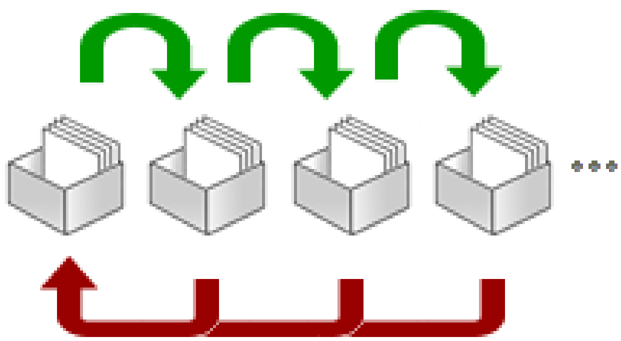

Figure 5.2: Flashcards sorted into levels. Arrows signify where a card is moved upon success or failure.

Leitner System The Leitner system is a very simple implementation of the concept. You have x levels, let's say 3 for this example. You put the things you know poorly in level 1, the things you know somewhat in level 2, and the things you know really well in level 3. You decide that you will rehearse level 1 every day, level 2 every week and level 3 every month. If your rehearsal of an item succeeds you move it up one level, if it fails you move it back to level 1. The number of levels and the intervals for each levels rehearsal is arbitrary.

Other solutions When you have a computer to do a lot of the work for you, the system for scheduling can be more complex. We can use a more dynamic algorithm that takes into account how long it's been since each rehearsal and how many of them you succeeded at. The result of this implementation creates intervals between an items rehearsal that can act more organic depending on how much you are struggling with that specific item.

Spaced repetition is not limited to flash cards though. As long as you have an item topic that you can quiz a user on, it should be possible to create a rehearsal schedule for it.

5.3 Recall

Recall is the retrieval of memories from the past. It is usually divided into: free recall, cued recall and serial recall.

Free recall is retrieval of information without any outside help. Cued recall is retrieval of information after first being giving a hint(memory from association). Serial recall is retrieval of a series of items(list of numbers, list of names, list of sounds...).

5.3.1 Staying focused

Henry Ford is known to have said: Nothing is particularly hard if you divide it into small jobs. He used this concept to make cars, but it can be used in many other ways as well. When doing big projects in work or school this paradigm of dividing big tasks into smaller subtasks is used often. There are many reasons for this.[13]

- It helps create an outline of the tasks to better understand the amount of work required.
- Helps you attack small tasks that are manageable instead of chewing on something bigger than you can comprehend.
- Easier to make sure there are no parts you have forgotten about.
- Makes it easier to see how far you've come towards completion.

All of these points make it easier to focus your work where it is needed and plan ahead. It also make the task less daunting as you don't have to digest the entirety of the project to do any work what so ever.

These are advantages that easily translate to the project of gaining new knowledge. When you are confronted with a curriculum or a language you want to learn, it is easy to be discouraged by this huge task. However, dividing tasks into manageable pieces will drastically help your ability to focus on the job ahead as well as lay a plan for future work.

Karl Kapp also mentioned this in his book "Gamification of learning and instruction"[11] Small achievable goals give a feeling of success which makes the student want to progress further. But it also help to be able to see your progress after the fact. Being able to compare your accomplishment to your ultimate goal, and see how you are doing compared to others. Then the student also get a competitive aspect helpful to some.

5.3.2 Saying the right thing

People's attention is limited, we need to make sure we spend it on what matters, or their attention will wander to things that seem more interesting. As said in an article by Gerhard Fischer:

"The challenge in an information-rich world (in which human attention is the most valuable and scarcest commodity) is not only to make information available to people at any time, at any place and in any form, but to reduce information overload by making information relevant to the task-at-hand and to the assumed background knowledge of the users."[19]

So why does the school system focus so much on students remembering every detail about their curriculum indefinitely when that is not a realistic goal and a waste of the students time? I would make a point of the fact that we only need to remember something for as long as that information is useful. When it is no longer valuable we can allow ourselves to forget it in favour of something else.

Learning is not an event, it is a process. The same can be said about forgetting. Even if you can not recall how to do linear algebra, having studied it in the past will make it easier for you to learn it again. Much easier in fact than keeping it in memory the entire time between you learned it at high school and you needed it again 10 years later.

You might think that some topics are vital parts in the core of mathematics or physics. That these topics need to be remembered in order to understand other subjects, we can't just let our students forget them. If something is such an integral part of the entire field of mathematics, you will have to use it regularly, and repeated recall is the key to long-term retention.[14]

5.4 user-interface

From the topics researched and the other teaching systems we can make some conclusions about what we want to achieve and how.

Just like in Khan academy, we will need to have two kinds of users. The students, and the administrators(teachers, teaching assistants, principals, researchers..). The main focus is to make sure the students have the best experience possible, but there are probably a lot of things we could do to benefit the other parties involved as well.

As discussed, any knowledge you learn will most likely be forgotten over time if not used regularly so setting a goal of remembering everything forever is illogical. That is why I choose to make this systems priority be to make sure they learn what they need to know when they need it.

Let's imagine a knowledge graph representing every topic in any subject you can think of, just like Gyldendal[4] wants to create a graph for mathematics. Maths are related to physics, chemistry as well, computer science too. We could make a graph for all these subjects interconnected by the nodes that are used by more than one of them.

Each topic has a list of other topics it depends on. This is represented by a line in the figure 5.3, and is the basis for how a student traverses from basic knowledge to advanced knowledge.

This helps you automatically divide anything you want to learn into smaller sub-goals which will help you work. Trusting that the system will schedule all the topics you need to know, you can focus on one thing at the time and not be distracted by all the other topics that would only make learning more difficult.

Choosing topics Students will be able to look through this tree of knowledge manually or search for a specific topic they want, or need to learn. They might have a test on "Adding binary numbers" for example, or their class has

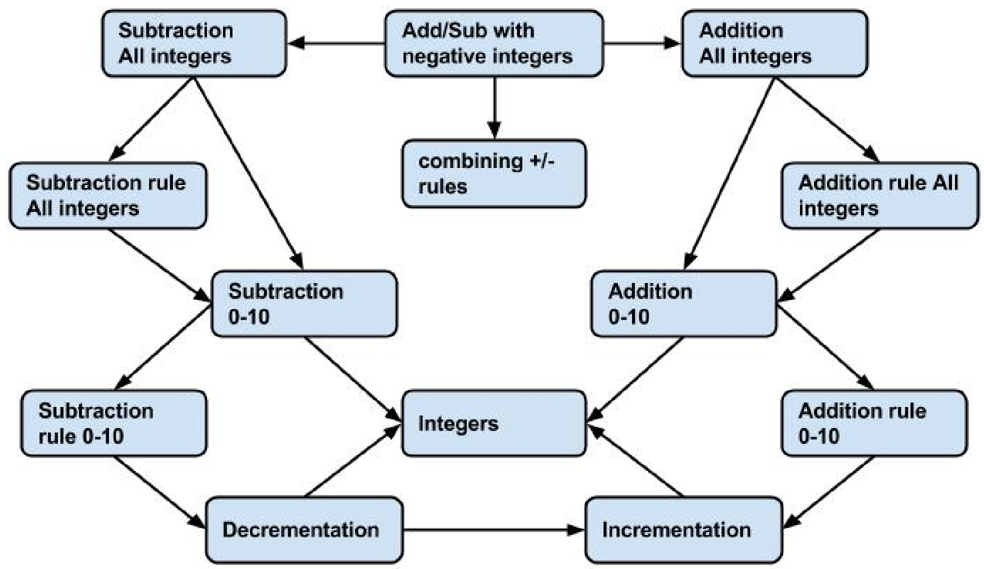

Figure 5.3: Example of a knowledge-graph around the topic of mathematics.

just started working on "Basic Algebra". "Adding binary numbers" would be one node in the tree. A piece of knowledge you want to learn. This node is then set as a target on your profile. "Basic Algebra" is a broader term, and encompasses many different nodes. Selecting this would therefore put every node related to it in the list of targets on your profile.

What to work on The main goal here is to make sure the students only has to work on topics they have control over, so the algorithm should go through each target node's dependencies and see if their level of mastery over those topics are high enough to warrant working on this topic now. If the system finds out they lack mastery over one or more of the dependencies it needs to check those nodes as well. Does these nodes have the necessary groundwork to warrant spending time working on them yet?

If the answer is yes, the student is tested by one or more relevant questions to see if she knows the topic better than anticipated. These questions must be generated randomly in order to ensure they remember how to solve them, not just the answer they gave last time. Having more than one question help against luck being a factor. This is similar to the mastery challenge in Khan academy where you get 5 randomly generated questions in a row and advance according to how well they answer. You want to make sure students are able to recall the information by themselves over time. This proves the knowledge is properly cemented in their long term memory. It also helps them remember it longer as it is repeated recall, not repeated studying that improves your retention over time.

Advantages Here the importance of how the topics are divided becomes clear. Just like how Codecademy only present very few new topics at the time, to keep the students focused on something small and manageable, each nodes topic should be small and manageable as well. This way you keep yourself from cluttering the short term memory with irrelevant information, making it easier to store transfer it to LTM. If something is complex the different items that makes it complex should be represented by that nodes dependencies. Then you are certain that the students working on this topic already know the different parts it is built on, so the only new thing they need to focus on is putting the different parts together. This way it is easier to keep focused when working and it is possible to do bite-sized advances to your learning without having to invest a lot of time on overhead. Having success when working is also a confidence boost who motivates you to work more. Dividing learning into small chunks therefore increases the amount of successes which in turn makes the students more productive.[11]

If they fail If the students are unable to recall the topic they will be provided with the tools they need to brush up on their skills. This could be a video explaining the topic, a well written article, a diagram, or all of these things. A version of the memorisation rules from Memrise could also be a good help. Some students might find one memory trick works really well while other students prefer something else. Letting them provide the rest with their own thoughts from a novices point of view, might help them together find a few good memorisation rules that is right for everyone.

Repeating challenges Because of the forgetting curve[20], the system needs to predict the likelihood of the students forgetting what they have learned earlier over time. This means they might be asked to do a challenge about a topic they have already finished, so that the system is sure you haven't forgotten it. This is a version of the repeated recall process. Every time they complete a topic's challenge, the system will take longer between each time that topic is tested. Eventually each test will be years apart, but they might not need that topic again for even longer and by then it's likely it has been forgotten.

I stress again that the goal of this system is not for the students to remember each topic forever. They only need to remember them for as long as they are needed. When a topic is no longer a target node, and no longer is a dependency of a target either the system will have no reason to refresh their knowledge about that topic until it becomes one again.

5.4.1 Other users

If you are a teacher using this system, you could make your own groups and choose what topics belong in it. A standard example is to put your class' curriculum in it's own group and have all your students choose that group as their targets instead of each student finding each topic themselves. If this group is private to the teacher and his students, every class can have their own group for their own curriculum represented in the system without any problem.

I can also imagine a teacher being able to group his students under himself. Letting him check each student's progress and see what topics they struggle with the most. This could be very useful for customizing what classroom activities they focus on. If he sees a student falling behind it's also easy for him to notice, so he can address the problem and give special attention where it is needed.

Just like all the students in a class is grouped under a teacher, every teacher can be grouped under a school principal, and every principal under a teaching district. This could let the principal see how well her students are doing and which teachers give the best results as well as the city or national educational institute can analyse the data for further study, and regulation of the curriculum. I see many applications for such a system to make education more manageable as a whole.

Saying all this does not mean the system can only be used in schools however. You could sign up as an individual and pick the topics you want to learn yourself. You could refresh something you learned through the system several years ago because you found out you needed it again. If you are a professor or researcher in the topic you could even help keep the system updated with the latest research and theories, letting anyone learn what you found out in the paper you wrote last summer. Or use it yourself to keep updated about what other researchers are doing. This all requires a fair bit of popularity to grow first of course.

Chapter 6

Using the cognitive architecture ACT-R as a basis for the adaptive user model

6.1 Building the graph

To represent knowledge of the universe through a graph, that graph needs to represent the universe as accurately as possible too. But the universe is complex, and even while only looking at a small piece of the puzzle it is hard to see which of the many permutations imaginable that is the most correct. Imagine 3 concepts. Addition, Subtraction, Multiplication. Are these concepts simple enough that we can represent them as a node?

Adding 0-10 Take addition. what do you need to know to add 3 + 2? You need knowledge of 3 and 2. Integers. You need to know what 3 is in relation to 2, and also 5. Counting, or incrementation might be a good word for that. 3 + 2 can be solved by a lot of different rules, but a simple one is: "3 incremented by 2" or "3 incremented two times". We now have a list of required knowledge. You must know what an integer is, you must know how to count(what incrementation is) and you must memorize this rule(x + y = x incremented y times).

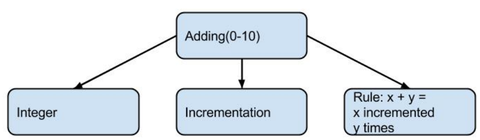

Figure 6.1: Figure of the knowledge graph surrounding the adding of smaller numbers.

Adding bigger numbers This works for all adding tasks involving integers, but what is the guy learning adding going to do when he get the task "10 + 134"? It is of course possible to count to 134, but doing so is not the easiest way to solve this problem. Here the most efficient solution would have to be to use a rule like the one you learn in school where you put the numbers under each other and add one row at the time. This is easier for bigger numbers, but more work than necessary for small numbers. So clearly some sort of division between the adding of big numbers and the adding of small numbers(0-10) is required. This new rule also relies on knowing how to add small numbers already. This is where the graph already begins to come together. Adding(0-10) requires: Integer, incrementation, rule(x + y = x incremented y times). Adding bigger numbers requires: Integer, incrementation, rule(placing numbers under each other and adding one row at the time.), Adding(0-10). The first requirements of adding bigger numbers already exists as a part of Adding(0-10) however, so we can simplify it to just the new rule and adding(0-10).

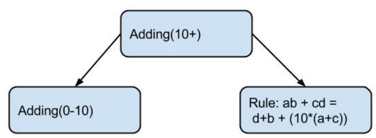

Figure 6.2: Figure of the knowledge graph surrounding the adding of bigger numbers.

However adding(0-10) is required knowledge to learn the new adding rule as well. And that the relation becomes more like this.

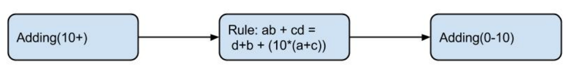

Figure 6.3: Alternate figure of the knowledge graph surrounding the adding of bigger numbers.

Alternative nodes There is still an issue however. This rule we have learned might not be the only rule to solve this problem, it might not even be the best one. Maybe we haven't come up with a better solution yet, or maybe the solutions whom are better are only better for yet another sub group of addition. For example adding of powers of 10, adding values of different units of measurement, or adding with another numerical system. Does being able to add any two numbers mean you have mastered addition if there are still new things you can learn to make your performance better? How do we distinguish between this.

As mentioned earlier we can only do our best to represent reality as accurately as possible, but this system should be modifiable enough to allow changes where needed in the future without compromising the data that has already been collected.

6.2 Cognitive architectures

Cognitive architectures are simply an attempt to create a model for the human mind and its psychology. The goal is to use it as the base of a computer programs (hopefully)intelligent behaviour.

John R. Anderson who did research on memory in the 1970's put together his theories on associative memory, thinking processes and long-term memory to create a cognitive architecture he called ACT. He used the term(cognitive architecture) as a name for the theories in his papers, since he didn't yet have any working implementation.

Because of this, cognitive architecture can refer to either the theory on how the mind is structured or an implementation of that theory on computers.[6]

6.3 The power law

The power law is defined as a relationship in which a relative change in one quantity gives rise to a proportional relative change in the other quantity, independent of the initial size of those quantities. A simple example of this is the relationship between the size of a square area and one of it's sides length. If you multiply a side by 2, you multiply the size by 4. Similarly if you multiply the side of a cube, you multiply the volume by 8. The factors don't depend on how large the square or cube is to begin with. Power laws highlight an underlying regularity in the properties of a system. So when we find them in natural or human phenomena that has great value to better understand them. An example is how the frequency of earthquakes varies inversely with their intensity, or how the number of people having a given income is inversely related to that incomes value.

The power law of practice states that the logarithm of the reaction time for a particular task decreases linearly with the logarithm of the number of practice trials taken.

6.4 Activation

Activation is a term that has been used in a lot of different fields, but in computer science it is used mainly as a tool in Artificial Intelligence to explain what parts of a person's memory is active based on a spreading algorithm. You start of with a network of nodes representing either abstract concepts in for example a neural network or factual concepts or information nodes. You define starting activation for the different nodes and then use weighted connections between the nodes to propagate activation between them. There are several ways to do this depending on how the network is built and what the nodes represent.

6.5 ACT-R

As described in the book "How can the human mind occur in the physical universe?"[3], there are several equations trying to simulate the brain and how it functions. Anderson and Schooler 1991 found that each time an item appeared, it added an increment to the odds that the item would appear again and that these increments all decayed away according to a power function. Thus if an item has occurred n times, its odds of appearing again is:

$$\sum_{k=1}^{n} At_k^{-d}$$

where t_k is the time since the kth occurrence. d is the decay parameter, usually defaulted as 0.5 and A is a scaling constant.

The natural logarithm of this value(called the base level activation) is then fed into an equation simply named the activation equation.

$$A_i = B_i + \sum_{j \in C} W_j S_{ji}$$

A_i is the activation of node i and B_i is the base level activation from the last equation. W_j is a weighting given from the connected node j and S_{ij} is the strength of that connection.

A_i(Activation of the node i) can now be fed into a new equation that returns the probability of recall.

$$P = \frac{1}{1+e^{-(A_i-\tau)/0.4}}$$

τ is a threshold set to -1 below which memories will not be retrieved.

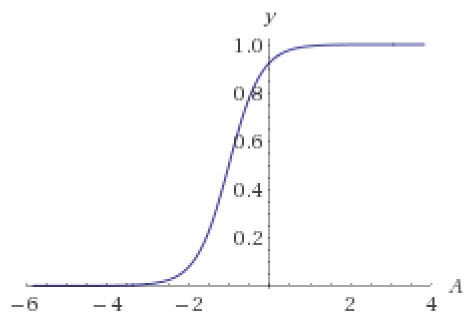

Figure 6.4: Graph of the probability equation.

6.6 User-model

To be able to track a students progress we need to make a model for that student. This model needs to be updated every time the student makes progress. But how do you model what someone know, or how well they know it?

In overlay modelling the user model is defined by a set of probabilities regarding how familiar the user is with a set of rules set by the expert. These probabilities are decided through logistic evidence about what rules the user has used when solving the problem in question, and what rules he has neglected.[2] One of the first uses of overlay modelling was to create a coach that can teach a user the game WUMPUS. It analysed the users moves, and used logic to find out if they had used the rules in the expert model. The system then changed how it coached them based on how much information it provided when providing lessons on what they had forgotten. If a basic rule already was being set as learned by the user, the system did not explain that rule when explaining a more advanced version of it. For example while telling you how to add big numbers, it would tell you to put the numbers above each other and to add each row one at the time. It would then also describe how to add numbers below 10 if the user didn't know that already in the user model. If the user did know it the system would not mention that information. Thus in this system the users are constantly being tested on their knowledge of the whole target topic, but the help they get when they make mistakes are customized by what basic knowledge they know already.

Because of the limited capacity of the working memory and the advantages to focusing on learning one thing at the time, I am not a huge fan of this. However the way users work themselves up towards learning something through subtopics and the system omitting to describe information it knows the user already knows are two aspects I also want to achieve.

The kind of system we want to create is one that builds on prior knowledge to ease learning of advanced topics. ACT-R describes a spreading of activation and the probability of retrieval of memory. We can use this to guide the user experience.

6.7 Implementing the activation spread

The base level activation described in 6.5 comes from timestamps for when that node has been recalled. Each time the users access a node for study, the system will have to note the time. Later it can calculate the time since the last access and feed it into the power function to see how much activation each earlier access point will contribute. This means that the more you access a node the more layers of inverse power functions lies on top of each other and the more time it takes for the sum of all activations to get low enough for another refresher on the topic to be necessary.

Connections Each node has different prerequisite nodes they are connected with. It makes sense that these nodes activation will have an impact on how well you know the node in question as well. If your knowledge of how to add smaller numbers and your knowledge of one of the rules for adding bigger numbers is solid, it will be very simple for you to use that knowledge in the node for adding bigger numbers. The second part of the activation equation reflects this behaviour. When calculating the activation of one node we simply have to go use that equation to improve our result. The W_j now refers to the activation of the prerequisite node j. I have no way of knowing which prerequisites are more important than others so I default S to one divided by the number of connections.

Probability threshold The activation will be a fluctuating value. Each time you refresh a node it will raise to a high value, and over time it will decrease. As the activation of a node decreases so does the systems guess at the users probability to recall it. At what point is the probability so low that we need to refresh it? Somewhere between 0.8 and 0.95 seems reasonable, but it's impossible for me to decide without experience in how it will effect the learning process. This is not something I have been able to test. Maybe it would even be advantageous to let the user change the value a little depending on what is best for them? If they have a very good memory they can lower the threshold and if their memory isn't as good they can raise it.

When the users make mistakes Most learning system with the use of spaced repetition give a heavy penalty when the user fails to recall the piece of knowledge that was asked from them. When using flashcards, cards the user fails to answer correctly on is usually put back to the beginning. Achieving this would be easy by just deleting all the timestamps the user has collected already. Theory however suggest that forgetting something does not mean it will be as hard to learn again as it was the first time, so deleting all the prior timestamps because of a failure seem to harsh.

My first instinct was to use a second variable named understanding to compete with the activation value. However I have not found any research or examples surrounding to support this. I can however make some suggestions. Because we want there to go some time between each refresh of a node we have to do something to the activation or it will remain below the probability threshold and then the system will just ask the user again right away. Adding a new time-stamp is also out of the question because that will make the delay until the next refresh too long.

One option is to move all the existing timestamps forward by the distance from the last one until now. This will reset the activation to the level and decay rate it had the last time the user refreshed the memory and succeeded.

6.8 Summary

We have a graph of knowledge.

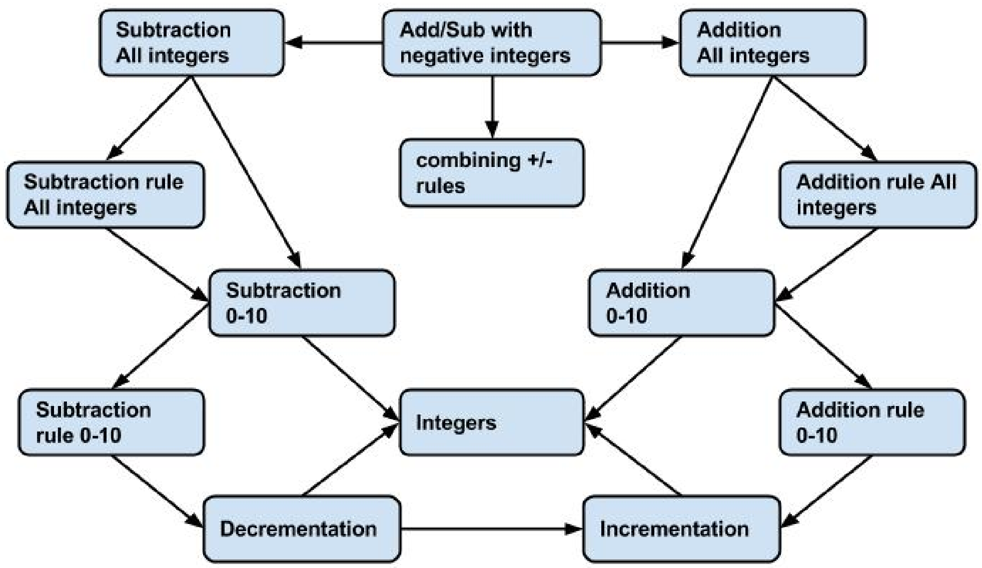

Figure 6.5: Refer to 5.3.

Each square is a topic to be learned. Each arrow points toward a dependence. Every time users work on one of the nodes that node gets a time-stamp. This time-stamp is used to calculate that nodes activation, a value that is then used to calculate the probability of the users recalling that information on prompt. A threshold is set so that when the probability of recalling a piece of knowledge that has been set as a target is too low, the system will help the user refresh it.

No node will be worked on unless that nodes dependencies have a high recall value. This makes sure that when trying to learn a complex topic they start with the basics, and work themselves up. However because the system has information on how well they remember things they have worked on before, some users will start off at a more advanced topic if their mastery of the basics are already sufficient.

The main issues are the creation of the graph and tweaking of variables. If we want to create a graph of all declarative knowledge, it will become immensely complex and prone to error. Not to mention that topics might need updates depending on new discoveries. Because of this it is important to have tools for maintenance. It needs to be possible to add or remove dependencies, or nodes when needed without destroying the data already collected on the users.

All the different variables also need to be tweaked in order to ensure the best user experience.

Another thing to consider is the people administrating the system. What kind of access should they have and what tools do they need to help the students they are responsible for.

Chapter 7

Evaluation

7.1 How can we create an adaptive learning system?

The first thing we have learned about learning is that we very easily forget. Whatever reason lies behind forgetting of memories, we need to build our system around this fact.

We need to make sure the users don't forget the things they need to remember. If a users really need to remember something, regular refreshing of the topics they are likely to forget will keep the memories fresh.

We need to make sure the users learn only what they are able to know with the knowledge they have, and don't assume they remember everything they have learned. By only giving them tasks they already remember the groundwork for, the users refresh some old knowledge while also progressing through the topics faster gaining a better understanding and motivation along the way.

We need to help the user stay focused. By only thinking about one problem at the time our brains are much more effective at learning. Guiding the user automatically through different topics, lets the user not have to worry about anything other than the task at hand

7.2 Using ACT-R as a basis for the user model

ACT-R uses time-stamps in order to produce activation values for different topics in an information graph. Each time stamp translate into a time interval that represent how much what you learned back then has decayed until now. Each time-stamp is then added together creating the base level activation of that information node. This activation is then used to calculate an estimation of the probability of that topic being recalled when prompted. This can be translated directly to the nodes in our knowledge graph.

Each time a user recall the information presented by one of the nodes it will save a time-stamp. Then when the system has to calculate weather or not a node needs to be refreshed it goes through the equations from the ACT-R model.

$$\sum_{k=1}^{n} At_k^{-d}$$

First it finds the decayed activation value from each of the time stamps (At_k^{-d}) and add them together.

$$A_i = B_i + \sum_{j \in C} W_j S_{ji}$$

Then this value is added to the signal provided by that nodes dependencies. Each dependency sends its activation and scale it by the number of connections.

Lastly the activation value provided by this equation is entered into a graph for calculating the probability of recall.

$$P = \frac{1}{1+e^{-(A_i-\tau)/0.4}}$$

In the end, this probability is used together with a threshold to decide what nodes need more work and not. It also helps decide if a node has the dependencies it needs to warrant studying it right away.

Chapter 8

Future-Work

8.1 Building knowledge graph automatically

A computer should be able to crawl the internet after information on different topics and dig out relations between them to create a sort of basis for the knowledge-graph we are after. The issues with this approach is the lack of organisation on pages we have no control over, and the use of different terms for the same thing on different pages. Some information might be false and some might be irrelevant. It is therefore important to have a way to clean up the graph after construction as well as moderating it. Add new parts without rebuilding it entirely and change/remove parts that might become outdated.

Semantic web would make this process a lot easier, but I don't see it as very likely that enough documents will have the information needed as well as note the information contained through the semantic web interface? Which is needed to make a good enough ontology.

8.2 Creating a knowledge base for easy sharing of new research in a topic

The first goal would be to map all the different pieces of knowledge that form the foundation of everything else. Mostly Elementary school and middle school curriculum. But as this process finishes mapping the high school curriculum and begin moving to University level subjects it would be interesting to see if this kind of knowledge map would be useful for those at the cutting edge of a field as well. Let a researcher add his research to the map of knowledge and add all the dependencies and related topics. He can then reference another researcher to this topic in the application and let him get an understanding of the groundwork the research is based on before he starts learning what this new research has found and why.

I also imagine that if someone had an idea they want to work on, it might be useful to be able to see what other work have been done on that topic. And a better overview on different topics might make this easier. Maybe you find someone has done some preliminary work already and you want to work with him to expand on what he has done. There's a lot of possibilities.

Another advantage to this is that you can see terms and labels others have put on aspects or phenomena or solutions, and easily see that they are talking about the same thing you called something else. This way you might easier create standard definitions for the things you describe in your research.

8.3 Support and tools for administrator

As mentioned the students would not be the only people to use the system. Teachers can use it to keep track of their students, principals can use to to keep track of their teachers. Someone need to provide expert knowledge to make sure the graph is correct and update it if it's needed. What kind of tools do they need? How do you make sure the user models are not destroyed in the process?

8.4 Creating a thread of topics to work on based on the graph generated and user knowledge estimated

As currently described the system only finds one topic at the time and chooses it arbitrarily among the possible candidates. It would be possible to rate them according to different parameters. Finding the most basic nodes first, finding the nodes you are closest to already knowing well first, or even finding the ones who needs the most work first.

Bibliography

[1] Atkinson, R. C., & Shiffrin, R. M. (1968). "Chapter: Human memory: A proposed system and its control processes". In Spence, K. W., & Spence, J. T. The psychology of learning and motivation (Volume 2). New York: Academic Press. pp. 89–195.

[2] Carr, Brian, and Ira P. Goldstein. Overlays: A theory of modelling for computer aided instruction. No. AI-M-406. MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB, 1977.

[3] Mellon, John R. Anderson Richard King. How can the human mind occur in the physical universe?. Oxford University Press, 2007.

[4] Bjørkeng, P. (2015). Datasystemet som holder elevene i flytsonen.http://www.aftenposten.no/fakta/innsikt/Datasystemet-som-holder-elevene-i-flytsonen-8010691.html Online; accessed 20-May-2015.

[5] Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A.Miyake & P. Shah (Eds.),Models of working memory(pp. 28 +mn 61). Cambridge, UK: Cambridge University Press.

[6] Anderson, John R. The architecture of cognition. Psychology Press, 2013.

[7] Thompson, Clive. "How Khan Academy is changing the rules of education." Wired Magazine 126 (2011).

[8] Tucker, Bill. "The flipped classroom." Education Next 12.1 (2012): 82-83.

[9] Bloom, Benjamin S. "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring." Educational researcher (1984): 4-16.

[10] Parkin, A. (1993). Memory: Phenomena, Experiment and Theory. Psychology Press Ltd.

[11] Kapp, Karl M. The gamification of learning and instruction: game-based methods and strategies for training and education. John Wiley & Sons, 2012.

[12] Tulving, E. and Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. Journal of Verbal Learning & Verbal Behavior, 5(4), 381-391.

[13] Altier, William J. The Thinking Manager's Toolbox: Effective Processes for Problem Solving and Decision Making: Effective Processes for Problem Solving and Decision Making. Oxford University Press, 1999.

[14] Karpicke, Jeffrey D., and Henry L. Roediger. "Repeated retrieval during learning is the key to long-term retention." Journal of Memory and Language 57.2 (2007): 151-162.

[15] Bell, Douglas S., et al. "Knowledge retention after an online tutorial: A randomized educational experiment among resident physicians." Journal of general internal medicine 23.8 (2008): 1164-1171.

[16] Bahrick, H. P., Bahrick, P. O., & Wittinger, R. P. (1975). Fifty years of memory for names and faces: a cross-sectional approach. Journal of Experimental Psychology: General, 104, 54-75.

[17] Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, 210, 207–209.

[18] Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory, (pp. 381–403). New York: Academic Press.

[19] 2001 Kluwer Academic Publishers. Printed in the Netherlands. Fischer, Gerhard. "User modeling in human–computer interaction." User modeling and user-adapted interaction 11.1-2 (2001): 65-86.

[20] Ebbinghaus H. Memory: A Contribution to Experimental Psychology. New York, NY: Dover Publications, Inc, 1964; original German text published 1885.

[21] McKenna SP, Glendon AI. Occupational first aid training: decay in cardiopulmonary resuscitation (CPR) skills. J Occup Psychol. 1985;58:109-117.

[22] Stahl, Stephen M., et al. "Play it Again: The Master Psychopharmacology Program as an Example of Interval Learning in Bite-Sized Portions." CNS spectrums 15.08 (2010): 491-504.