Information Theory on Czech Grammar Schools: First Findings

Daniel Lessner
Department of Software and Computer Science Education,
Faculty of Mathematics and Physics,
Prague, Czech Republic
lessner@ksvi.mff.cuni.cz

ABSTRACT

Computing science is not a part of standard grammar school (secondary) education in the Czech Republic. As a part of an effort to change this, we have developed an introductory computing science course.

It consists of multiple month long modules. The first module deals with the notion of information. Students are expected to understand it sufficiently for comparing informational gain of statements or questions. Along with binary numeral system, they are expected to understand the relation between data length and the amount of information stored.

In this paper we describe the module itself and present the results from the first run of experimental teaching. The aim was to find out the limits of such topic on grammar schools and suitability of chosen teaching methods.

It turned out that no matter how abstract is the term information itself, students are capable of acquiring it deep enough so that it would be beneficial for them.

Categories and Subject Descriptors

K.3.2 [**Computers and education**]: Computer and Information Science Education—*Curriculum, computer science education*

General Terms

Human factors, Experimentation

Keywords

Secondary education, mandatory foundations, information theory

1. INTRODUCTION

In this paper we first briefly discuss the unfortunate situation of computing science (CS) education on Czech grammar schools[1]. Then we describe a basic CS course developed with an intent to change this state. This course is tested with two seminar groups. The aim of this experimental teaching is to find out whether it is even possible to teach CS on a grammar school level in Czechia.

We have decided to focus this paper on the first module, which helps students study information. We describe its specific objectives and content. The necessary aspects such as the course organization and students characteristics are described in a separate section. The last section finally shows the results and findings and their interpretation.

Grammar schools in the Czech Republic are one of the secondary education branches. The curriculum is defined in the Framework Education Programme for Secondary General Education (FEP) [1]. From it the individual school programmes are derived. According to FEP, grammar schools shall prepare students for university studies. This shall be accomplished through development of key competences. FEP defines six of them: to solve problems, to communicate, to learn, civic, entrepreneurial and personal and social competence. Classical subjects are means to develop these key competences[2].

We use the term computing science for the general area of efficient (i.e. automated) information processing. It has very little to do with digital literacy. It is also not programming. Computer is merely the tool, not the subject. A similar view can be seen in Computer Science Unplugged from New Zealand [2]. It does not employ computers by its definition, what certainly says enough about its relation to digital literacy or programming.

The K-12 CS curriculum is not widely applied [5], but serves as a good prototype and reference. Given a much longer time frame, the programme can afford to be much more complex then the other examples here. However, its goals are similar to ours: to introduce basics of CS on an appropriate level. All topics included in our course except for a few details are covered in this curriculum.

FEP casually mentions a few CS topics. It includes so called informatics[3] as a compulsory subject, which is focused on using digital technology. There is no systematic approach to efficient information processing.However, considering the

[1]General secondary education, students aged 15 – 19.
[2]Yet they are not unimportant – to communicate we need languages, to solve problem sciences, to be good citizen history etc.
[3]Informatics in Czech is an umbrella term, which can mean anything related to computers or digital technology, depending on context.

WIPSCE '12, November 8–9, 2012, Hamburg, Germany.

grammar school educational goals and the concept of key competences, CS seems to be very promising. Moreover, it is a part of general knowledge [3]. CS is shaping our world significantly and most of grammar school graduates are not even aware of its existence.

We believe we are missing here a fruitful opportunity to achieve secondary education goals easier. In an imaginary ideal case, CS is a regular school subject, such as physics or history. We intend to develop, carry out and evaluate a basic CS course to have a solid base for a possible future discussion about including CS into grammar schools.

This paper presents results from one course module. The question is, whether it is possible to teach information and to what extent. We have planned the module (objectives, content, activities, assignments, final test) and carried it out with two seminar groups.

2. INTRODUCTORY CS COURSE

In an effort to move towards our above mentioned ideas, we have developed a program for a basic CS course. It covers what we think any grammar school student should know. General goals of the course are derived from secondary education goals defined in FEP. We shall enhance students problem solving skills along with other key competences and thus prepare students for studying university, for their career and for a life in our society [1].

Further, as with other sciences, the aim of their study on grammar schools is to get familiar with their subjects, methods, connections to other areas, fundamental results and their implications and applications. Specifically for CS, students shall be aware of some fundamental limits it has, such as the existence of practically or computationally unsolvable problems. Students shall be able to actually utilize the basics of CS in everyday life.

Based on the usual situation on our grammar schools, we have planned the course to last one school year, 90 minutes weekly. It is structured into approximately one month (four lessons) lasting modules. Each module deals explicitly with one fundamental idea, while other ideas are present in the background. The main topics chosen for individual modules are information, graph, problem, algorithm and efficiency. More detailed information can be found in [4].

3. INFORMATION: MODULE PLAN

Information is the first module out of 8 in the course plan. The word informatics used in Czech is based on the word information, so it seems natural to deal with it in the beginning of the course. But most importantly, this module contains many notions that will come later (such as algorithm, efficiency or recursion) in a simple and conceivable form.

3.1 Objectives

The basic objective to achieve is understanding of the term information on an appropriate level and understanding of what is it good for. Along with the term information itself, there are others to be understood: encoding, bit, data, message, communication, language, knowledge. Further, students shall be able to calculate the amount of information contained in a message.

By a step higher is efficient querying in a given situation. This goes along with the concept of halving the set of possibilities. Students shall understand the concept of decision tree, use it properly, judge its efficiency and optimize it in simple cases (including non-uniform distribution of answers). Students shall be able to work correctly with the relation between the number of codes and their length.

As the CS course opens with this module, we use it to prepare and develop concepts for upcoming topics, such as decomposition, algorithm (a reliable workflow of simple steps, e.g. a decision tree) or efficiency (related to defined criteria).

3.2 Process

Here we describe briefly the educational activities and related content included in the module. They result from the above discussed objectives and methods. There is some slack to vary the activities between each study group.

Students are at first introduced to the course and to the content and objectives of the first module. Then the teacher helps them realize the importance of the term information and the fact that they actually can not explain what it is. They are asked to search their resources and come up with an explanation themselves. Discussing their various results and opinions, key features and links emerge.

Guessing games

To make things less abstract and vague, the teacher proposes an animal guessing game. He thinks of himself as of an animal, students ask him yes or no questions in an effort to determine the animal as fast as possible. Students intuitively review the efficiency of each question after the game and formulate their first hypotheses on the optimal strategy. Frequent finding is for example that direct guessing is useless in the beginning, negated "reassuring" question is useless completely in noise-free environment, making the teacher answer *yes* may be emotionally satisfying, though it is meaningless for the game.

However, animals turn out to be a rather too complex model. They are not a well bounded domain, success in the game depends on some knowledge of animals, and it is surprisingly tricky to answer some questions (e.g. elephants do live in Africa, as well as in India, and in captivity all over the world).

That is why we switch to a much better arranged domain – natural numbers. They still *model* the situation well enough to give us desirable results. Numbers allow students to clear their minds from animals and focus on the problem itself. The task is to formulate a good strategy, i.e. find a set of rules to create efficient questions. Some heuristics emerge eventually in most of the pairs. Then we play one more game together so that each pair can see their results in practice.

Decision trees and information definitions

One of the ways to describe a strategy is a decision tree scheme. It shows the efficiency also visually. Looking at a tree scheme, one may work out the formula to the relate number of possibilities and and number of questions asked to isolate one. This finally allow us to quantify information. Considering the scheme, it is practical to measure uncertainty by number of questions necessary to remove it, i.e. $\lceil log(|S|) \rceil$, where S denotes the set of possible outcomes[4]. Base of the logarithm is two in case of binary questions.

[4]If students do not know logarithm yet, they simply emulate it by iterated division.

Amount of information is then the decrease of such uncertainty after getting the answer. More formally, the amount of information in a message M, represented as a set of eliminated outcomes, is $I(M) = log(|S|) - log(|S \setminus M|)$. This approach is a special case of the well known approach using entropy, using uniform probability distribution. This makes the basic idea of *reduction of uncertainty* more accessible as *reduction of possible outcomes.* This simplified definition is sufficiently useful to lead to the main ideas behind.

The second definition of information we use says that information consists of data and their interpretation. This is a fruitful starting point to further observations too. It is in accordance with students intuitive concepts. It can be illustrated using the decision tree scheme too. Another closer look to the decision tree for guessing numbers reveals the tight connection to binary numbers. It also opens the opportunity to think of an informational value of an individual character, e.g. a bit.

Students realize sooner or later that not every animal has the same chance to take part in the game. The teacher provides them with some more examples where non-uniform distribution is important, e.g. assessing stroke risk factor or other medical diagnosis tasks. This topic leands to a fundamental heuristic: what comes often shall be at hand.

Additional activities

Most of the objectives of this module are covered with activities above. Below suggested activities are optional and utilized mostly based on students interests. They serve to review the matter and to investigate it from different points of view. They also imply some real world applications. Not all students in the group must work on the same task of these.

The guessing game can be extended for unlimited interval, rational and real numbers. It of course deepens the understanding of mathematics. For CS, it opens the question of a process finiteness with an couterintuitive twist: an infinite number of objects to guess does not necessarily cause infinitenes of the inquiry process.

Another domain for guessing game may be subsets of a given finite set. Halving the possibilities leads to the simplest encoding by characteristic function. And again, students reach deeper understanding of sets and subsets, which will be useful later in the course. Another opportunity to use the matter is to play mastermind.

To remind students that binary questions are just a convention, we let them solve problems with balance scale. One kind of problem is to find a heavier item in otherwise indistinguishable group. Another kind is about designing somehow optimal set of masses.

Popular TV shows, such as *Who Wants to Be a Millionaire?* can be a good resource. Students can think of heuristics on which kind of help shall be used in which situation. It is easier to explain these heuristics using the new knowledge of information theory. A little bit similar and very openly defined project deals with designing cheat sheets, where efficient encoding is of course crucial.

Efficiency of information encoding in natural languages can be investigated by comparing same texts in different languages. Another question of this field is the efficiency of Morse code for different languages, and to find a way to enhance it, if possible. Also DNA encoding amino acids is an interesting system to be investigated from the informational point of view. The last one to mention is the START method (Simple triage and rapid treatment) utilized in mass disasters.

3.3 Feedback and evaluation

Students work mostly in groups, what allows them to receive continuous feedback on their thoughts from their peers as well as from other groups. They also consult the teacher when needed during their work, what allows him to provide some more informal feedback. Another source of feedback are submitted home assignments, commented by the teacher. They are usually problematic or require some more intensive work, which can be done outside the classroom.

Students write a test at the end of each module. These tests are rather difficult. The matter itself is quite abstract, the tests last the whole lesson (90 minutes) and contain some unusual tasks. Some aspects on the other hand make the tests easier. The required skills are stated very clearly in advance. Further, students may use their lecture notes and internet resources.

4. EXPERIMENTAL LESSONS

Here we describe the actual teaching to allow the reader to better understand the below presented results. We worked with two groups. Both groups are of 12 students. Majority of students does not intend to continue their carrier in CS anyhow. Their motivation is only to pass the subject as effortlessly as possible. Almost a half of them have signed up for the seminar as for "the best of bad options“, because they just had to pick one. This fact of course hinders teaching CS. On the other hand, it increases the validity of our research. Our target is a regular grammar school student. The less interested students (i.e. closer to the average) we work with, the more valuable results we obtain.

All twelve students from the first group are in their last year. Their school specializes on sports. This determines their attitude to sciences well enough (and again, it increases the validity of our results). They all have been through a one year lasting optional seminar on programming in Pascal.

Four students of the second group are in their last year and they all intend to pass the school-leaving exam in informatics. The others (eight students) are one year younger and their decision is still to be made. The students have very various programming experience.

The module on information took six weeks (90 minutes teaching time each) instead of the originally planned four. This is mostly due to building in the necessary mathematics reviews (e.g. numeral systems and logarithms).

5. RESULTS

Students accepted the module surprisingly well (considering it being quite deep and abstract and out of the official curriculum). Some of them realized the importance of the topic and were glad to understand it better, some were challenged by the proposed problems.

5.1 Observation during lessons

An unexpectedly fruitful discussion arose from "Is it even?“ question during numbers guessing. It is informatically optimal. Some students therefore insist to use it as the first question. Some realize that they can ask it also at the end in a simpler way, when they are choosing between the last two alternatives. This shows an important fact: difficulty of answers management is important in practice. The question does not allow us to construct consequent questions straightforwardly with the same principle. Yet, after being reminded of binary system, students are able to realize the connection. They are doing almost the same as bisecting, only reversed. The value of these thoughts lay in the fact that they were driven mostly by students themselves.

Disturbing (however expected) finding is low level of understanding of certain parts of mathematics among students. Logarithms are taught intensively, yet no (!) student have seen a direction to calculate the depth of a tree. Our guidance had to be very explicit. A few months later, most of the students actively knew that it is related to exponential growth and therefore also to logarithms.

Some of the ideas encountered in the first module were needed again during the school year. More than a half of students (in both groups) still recalled them actively. This includes bisection, decision trees, binary numeral system and efficiency modeled using some kind of elementary step.

5.2 Test results

Further explanations and remarks are based on collected test sheets examination and our own notes from the lessons and consecutive discussions with students. We have not used any formal pre-test. It was more than clear from the initial discussions held for this purpose that students had difficulties to even understand the given questions.

Brief explanation.

Students were to explain briefly the meaning of the term information. Definition or equally complete description got all the points, less precise expressions or examples got only partial score. Any meaningful answer was rewarded with some points. The resulting success is lower than expected, given that it is a question of knowledge and outer resources were available. We see the reason in the lack of communication skills. Most of the students understood the term well, as we have seen during other tasks. To express it was however too difficult.

Information and logarithm.

This was a double task. First was to determine how many yes/no questions does one need to reliably identify a Czech deputy (they are 200). The second was to determine the minimal sufficient length for their hypothetical binary ID code. Very few students have used logarithm for any of the tasks. Majority of them chose to simulate the known process (sequential halving and doubling, respectively. Only two students have realized that the result is the same for both questions and the second calculation is not needed.

Suboptimal decision tree.

There was a small (7 leaves, 4 levels) decision tree with given rates per leaf. First task was to calculate the expected number of questions asked. The second was to lower that number by adjusting the tree.Those who made some calculations usually got both parts right. Less than a quarter of students (in both groups) passed the calculation, as it was perhaps still too mysterious for them. On the other hand, some of them understood the situation well enough to optimize the tree (swapping frequent nodes towards the root).

Inquiries comparison.

Students were to sort seven given questions (such as *"Is it a prime?"*) according to the expected amount of information provided by the answers. All the questions were bound to the same situation, identifying an integer number between 0 and 255. This task was very dependent on mathematical skills, therefore the results are rather poor.

Statements comparison.

This last task was not included in the same test with the others. We put it into another test six months later. There were four different smiley-like crime suspect faces and four statements regarding the offender, such as *"The offender has no hair"*. The task was to sort these statements according to the amount of information included. Unlike with the previous task, here it was obvious enough that those who had a mistake in their results did not understand the matter at all. Hence we gave all the points or none. The success rates were unexpectedly high, considering the time passed. We can conclude that the educational results hold well.

6. CONCLUSION

We described the concept of an introductory CS course for Czech grammar schools and gave more detailed specification about the module about information. Then we could proceed to the results obtained during the first testing year of the course. The most distinctive feature of the proposed course is the shift in goals from CS itself to key competences and connections to other grammar school subjects.

Based on the results above, we may conclude that students are capable of understanding the notion of information and its basic applications. The goals of the module were set surprisingly close to the real need. The limiting factors lay mostly in students mathematics skills.

The teaching methods are found useful and efficient by the students, although they have recommended some tweaks, mostly in time allocation for each activity. They are also uncomfortable with open problems and want to know specifically what to do as soon as possible[5]. We have to incorporate more frequent formal evaluation. The weaker students seem to be rather insensitive towards the informal or peer feedback.

7. REFERENCES

[1] *Framework Education Programme for Secondary General Education (Grammar Schools)*. Výzkumný ústav pedagogický v Praze, 2007.

[2] M. R. Fellows, T. Bell, and I. Witten. *Computer Science Unplugged... offline activities and games for all ages: Original Activities Book*. Computer Science Unplugged, 1996.

[3] J. Hromkovič and B. Steffen. Why teaching informatics in schools is as important as teaching mathematics and natural sciences. In *Proceedings of ISSEP*, ISSEP'11, pages 21–30, Berlin, Heidelberg, 2011. Springer-Verlag.

[4] D. Lessner. Computer science curriculum proposal for czech grammar schools. In *Zborník príspevkov*, ITAT 2011, pages 99–104, Seňa, Slovakia, 2011. PONT s. r. o.

[5] A. Verno. Supporting k-12 computer science education. *J. Comput. Sci. Coll.*, 23(3):145–146, jan 2008.

[5] This effect somewhat faded in the latter modules.