

Replace the contents of this file with official assignment.
Místo tohoto souboru sem patří list se zadáním závěrečné práce.

Bachelor's Thesis

FINITE AUTOMATA IPAD EDITOR

Marek Fořt

Czech Technical University, Faculty of Information Technology
Department of Theoretical Informatics
Supervisor: Ing. Jiří Trávníček, Ph.D.
April 13, 2021

Czech Technical University in Prague
Faculty of Information Technology

© 2021 Marek Fořt.

This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).

Citation of this thesis: Marek Fořt. *Finite Automata iPad Editor*. Bachelor's Thesis. Czech Technical University, Faculty of Information Technology, 2021.

Contents

Acknowledgment	v
Abstract	vi
1 Introduction	1
1.1 Motivation, Focus of Thesis	1
1.2 Thesis Goals	1
1.3 Thesis Structure	1
2 Theory	3
2.1 Formal Languages and Grammars	3
2.2 Finite Automata	5
2.3 Machine Learning	7
3 Analysis	9
3.1 Existing Applications	9
3.2 ALT Analysis	12
3.3 FA Model	12
4 Automata Editor Design	13
5 Implementation	15
6 User Testing	17
7 Conclusion	19
7.1 Goals Assessment	19
7.2 Thesis Contribution	19
7.3 Future Work	19
8 List of Abbreviations	21
A Nějaká příloha	23
Obsah přiloženého média	27

List of Figures

2.1	FA graph representation	6
2.2	FA table representation	6
3.1	ALT web interface screenshot	10
3.2	Statemaker screenshot	11
3.3	TuringSim interface screenshot	11

List of Tables

Seznam výpisů kódu

Acknowledgment

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis. I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as a school work under the provisions of Article 60(1) of the Act.

In Prague on 13th May 2021

.....

Abstract

Write abstract here!

Keywords iPad application, finite automata, interactive editor, AlgorithmsLibrary Toolkit, Composable Architecture, Swift

[illegible]

Introduction

1.1 Motivation, Focus of Thesis

Theory of finite automata is an important part of computer science curriculum at FIT CTU in Prague and other universities around the world. And although there is a lot of resources one can learn from, there is a lack of those that utilize modern tools. One of such modern tools is iPad (and touch devices in general). This thesis will fill in this gap as the result will be a finite automata editor application for iPad.

Furthermore, I will expand on the recent work done at FIT CTU in Prague concerning development of algorithms library and, more importantly for this thesis, finite automata algorithms including simulating input. This library is named Algorithms Library Toolkit (ALT) [14] and it has been open sourced.

The main motivation of this thesis is to improve how students learn finite automata and more specifically, enhance the current course BI-AAG that is taught at FIT CTU in Prague. It is also an opportunity to try out algorithms library in practice and create a concrete example of how it can be leveraged.

1.2 Thesis Goals

The main goal of this thesis is to implement a prototype of an automata editor for iPad. This application should enable users to create and edit finite automata with emphasis on touch-based input. I will also study the web interface of ALT [13] [16], ALT itself focusing on design and drawing of finite automata, the possibilities of strokes detection on touch devices, and the approaches of shape detection, especially those used in automata drawing on iOS platform.

After the initial study of current approaches and theory, I will implement a prototype of a finite automata editor iPad app that will be capable of recognizing finite automaton elements from strokes and simulating input.

I will then conduct usability testing to assess the usability and shortcomings of the prototype.

1.3 Thesis Structure

Let me now introduce you to the structure of the rest of the thesis:

- In **Chapter 2** I will go over the theoretical concepts to properly explain terms and concepts on which it will be built upon later.

- **Chapter 3** is concerned with the analysis of already existing solutions of creating automata editor, the existing ALT web interface and ALT itself.
- **Chapter 4** is about the design of the editor itself.
- In **Chapter 5** I will write about the implementation.
- **Chapter 6** will go into the specifics of user testing and its outcomes.
- **Conclusion** is the last chapter of this thesis where I will assess the success of fulfilling aforementioned goals and lay out possible future development.

Chapter 2

Theory

Firstly, I will need to define terms and formal definitions concerning mainly finite automata theory, as that is the main subject of this thesis, and then machine learning as some of its concepts were important during the implementation.

2.1 Formal Languages and Grammars

The following definitions are taken from Automata and Grammars by Eliška Šestáková [10], Introduction to Automata Theory, Languages, and Computation [6], and materials from BIE-AAG course [5].

2.1.1 Formal Languages

► **Definition 2.1.** *Alphabet (conventionally denoted by Σ) is a finite set whose elements are called symbols.*

Alphabets therefore can be:

- $\Sigma = \{0, 1\}$
- $\Sigma = \{a, b, c, d, e\}$
- $\Sigma = \{\text{one, two}\}$

► **Definition 2.2.** *String (word) over an alphabet is a finite sequence of symbols from that alphabet.*

- ϵ - empty string (string with zero occurrences of symbols)
- Σ^* - set of all strings over Σ
- Σ^+ - set of all nonempty strings over Σ

For a binary alphabet $\Sigma = \{0, 1\}$ $\epsilon, 1001, 100, 1, 001$ are all strings over the alphabet Σ .

► **Definition 2.3.** *Formal language L over an alphabet Σ is any subset of all the strings over Σ - i.e., $L \subseteq \Sigma^*$*

For a binary alphabet $\Sigma = \{0, 1\}$ a formal language over Σ is then subsets of *all* binary strings. We can denote the language either by:

- enumeration notation where all possible strings in the language are listed, e.g.: $L_1 = \{\epsilon\}$, $L_2 = \{1\}$, $L_3 = \{0, 00, 000, 01\}$.
- set-builder notation where the languages are described in the following way: $\{w \mid \text{something about } w\}$. Examples are: $L_4 = \{w \mid w \in 0, 1^* \wedge |w| \bmod 2 = 0\}$, $L_5 = \{0^n 1^n : n \in \mathbb{N}_0\}$.

2.1.2 Grammar

Grammars are used to describe languages. Below you can find how they are defined:

► **Definition 2.4.** *Grammar is a quadruple of $G = (N, \Sigma, P, S)$ where:*

- N is a finite non-empty set of nonterminal symbols.
- Σ is a finite set of terminal symbols ($\Sigma \cap N = \emptyset$). Note that $N \cap \Sigma = \emptyset$.
- P is a finite set of *production rules*, assuming the following form:

$$\alpha A \beta \rightarrow \gamma \quad (\alpha, \beta, \gamma \in (N \cup \Sigma)^*)$$

The following is an example of a grammar that describes the language $L = \{01^n 0 : n \in \mathbb{N}_0\}$: Grammar $G = (\{A, S\}, \{0, 1\}, P, S)$ where P :

- $S \rightarrow 0A$
- $A \rightarrow 1A$
- $A \rightarrow 0$

2.1.3 Chomsky Classification of Grammars

Grammars are divided into four classes where they differ in their production rules.

► **Definition 2.5.** *Let $G = (N, \Sigma, P, S)$. We say that G is:*

1. *Unrestricted grammar* (type 0), if every rule is in the form of:

$$\alpha A \beta \rightarrow \gamma \quad (\alpha, \beta, \gamma \in (N \cup \Sigma)^*, A \in N)$$

2. *Context-sensitive* (type 1), if every rule is in the form of:

$$\gamma A \delta \rightarrow \gamma \alpha \delta \quad (\gamma, \delta \in (N \cup \Sigma)^*, a \in (N \cup \Sigma)^+)$$

or in the form of $S \rightarrow \epsilon$ if S is not present on the right hand side of any rule of a given grammar.

3. *Context-free grammar* (type 2) if every rule is in the form of:

$$A \rightarrow \alpha \quad (A \in N, \alpha \in (N \cup \Sigma)^*)$$

4. *Regular grammar* (type 3), if every rule is in the form of:

$$A \rightarrow a \text{ or } A \rightarrow aB \quad (a \in \Sigma, A, B \in N)$$

or in the form of $S \rightarrow \epsilon$ if S is not present on the right hand side of any rule of a given grammar.

2.1.4 Classification of Languages

Classification of languages, also known as the Chomsky hierarchy, has the following definition:

► **Definition 2.6.** *We say that language is:*

1. *formal* if it is a formal language but is neither regular, context-free, context-sensitive, nor recursive enumerable. These languages are not recognized by Turing machine.
2. *recursively enumerable* if and only if \exists unrestricted grammar which generates it
 - recognized by Turing machine
3. *context-sensitive* if and only if \exists context-sensitive grammar which generates it
 - recognized by linear bounded Turing machine
4. *context-free* if and only if \exists context-free grammar which generates it
 - recognized by a nondeterministic pushdown automaton
5. *regular* if and only if \exists regular grammar that generates it
 - recognized by finite automaton

In this thesis we will be mainly interested in regular languages / grammars since those are recognized by finite automata. Finite automata will be defined in the following section.

2.2 Finite Automata

The final editor app will be for finite automata, therefore they are very important for this thesis. Informally, a finite automaton is a model for simple computation. States, that serve as memory, and transitions together form a *control unit*. Along with a control unit, the finite automaton has a *read-only input tape*, which is divided into individual cells, and the *head* that scans the input tape as the automaton continuously reads it, cell by cell. Automaton starts in its initial state and with head pointing at the first cell. As the input is read, the head moves until it has read all of the input tape. If there is a missing transition for an input, the automaton does not accept the input. Otherwise, it accepts the input if it is in an end state at the end of the input.

Let's define a finite automaton formally:

► **Definition 2.7.** *Finite automaton is a quintuple $M = (Q, \Sigma, \delta, q_0, F)$ where:*

- Q is a finite non-empty set of states
- Σ is a finite input alphabet
- δ is the transition function (the exact definition is determined by which type of finite automaton it is - see below)
- $q_0 \in Q$ is the initial state
- $F \subseteq Q$ is the set of initial states

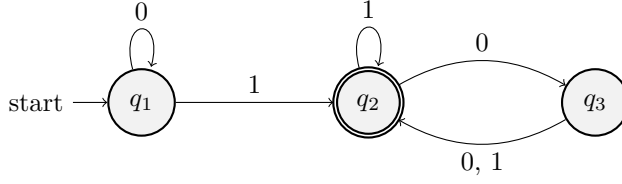
Finite automaton can also be either *deterministic* finite automaton or *nondeterministic* finite automaton. This dictates the exact definition of δ - transition function. For deterministic finite automaton (DFA) the definition of δ is:

δ is a mapping from $Q \times \Sigma$ to Q

δ for nondeterministic finite automaton (NFA) is defined as:

δ is a mapping from $Q \times \Sigma$ into the set of all subsets Q (denoted by 2^Q)

Expanding upon the difference between the definition of the transition function:



■ **Figure 2.1** FA graph representation

δ_{NFA}	0	1
$\rightarrow S$	S	S, A
A	B	
$\leftarrow B$		

■ **Figure 2.2** FA table representation

- DFA can only transition from one state to another, e.g. from q_0 to q_1 ($q_0, q_1 \in Q$)
- NFA can transition to a set of states, e.g. from q_0 to q_1 ($q_0, q_1, q_2 \in Q$)

If we change the definition of NFA's δ to a mapping from $Q \times (\Sigma \cup \{\epsilon\})$ we allow, what are called, ϵ -transitions that allow us to move to a different state while not reading any input from the tape. This finite automaton is then called *nondeterministic finite automaton with ϵ - transitions*.

2.2.1 Representation of Finite Automata

Finite automata's transition functions δ are generally represented in the form of:

- *Formal notation*
 (NFA) $\delta(S, 0) = \{S, A\}$ (transition from the state S and symbol 0 to the states S and A)
 (DFA) $\delta(A, 0) = B$ (transition from the state A and symbol 0 to a single state, not a set of states, B)
- *Weighted directed graph* (state diagram)
 Automata can be represented graphically as directed weighted graphs. Each state is represented as a vertex in the graph and final states are recognized by being a double circle, instead of a single one. Initial state is the one with an incoming edge. The transitions are then directed edges between states. You can see FA represented as weighted directed graph in 2.1
- *Table*
 Table representation has in the first column all states where initial state is marked with \rightarrow while final states is marked with \leftarrow . In the first row, excluding the first column, there are symbols of the alphabet, Σ . In the rest of the rows are states (or a set of states) that will be transitioned to on a given input (defined in the first row). You can see an example of it in 2.2.

In this thesis we will mostly be working with the representation in form of weighted directed graph as that is what will the user edit in the app. This also concludes theory about finite automata and formal languages.

2.3 Machine Learning

Machine learning does not have an exact definition but e.g. in a book Foundations of Machine Learning it's loosely defined as "computational methods using experience to improve performance or to make accurate predictions" [8]. *Experience* means something we know from the past that we can leverage for making predictions in the future. Usually, this experience comes in the form of data. The book Foundations of Machine Learning [8] and materials from BIE-VZD from FIT CTU in Prague [3] will be used further in this section to define terms and concepts necessary for this thesis.

2.3.1 Classification

Machine learning, in order to cluster problems that can be solved in a similar way, have defined a few learning scenarios, most notably supervised and unsupervised learning. Learning scenario is a basic description of what type of data we have, how we receive the data and the test data that we use to evaluate the learning algorithm.

- *supervised learning*: Our goal is to explain *variable* Y given *independent variables* X_0, X_1, \dots, X_{p-1} . We do this by finding a "function" for which most of its examples the following holds:
$$Y \approx f(X_0, X_1, \dots, X_{p-1})$$
- *unsupervised learning*: Our goal is to find structures of "similar" data. We do not predict any class and there is no clear way to assess the quality of an unsupervised learning algorithm since it is not clearly defined what the end result should be.

In this thesis we will be only interested in the supervised learning. We can also divide common problems that machine learning is trying to solve by learning tasks - that includes classification, regression, ranking, clustering, etc. Let's look more closely at classification which will be later used in the implementation.

- Classification is a problem of assigning a category to each item.

It is also a problem solved via supervised learning. To expand on the definition of supervised learning from above, classification is a special case where Y has only a few (countable amount) of values. The simplest example of classification is *binary classification*. E.g. we want to predict whether a patient has flu and our data - gender of a patient, person can leave the bed - can be represented in a binary format (yes/no).

Chapter 3

Analysis

In the analysis I will study the following:

- existing applications that enable users edit finite automata
- ALT itself, focusing on design and drawing of finite automata
- possibilities of detection of strokes on touch devices.

3.1 Existing Applications

This section will be concerned with the study of existing applications - be it applications for mobile or web.

3.1.1 ALT Web Interface

ALT web interface has been built as a part of bachelor's thesis made by Michael Vrána [16] leveraging work already done in ALT itself. ALT web interface uses Pipe-and-Filter [2] architecture to easily combine input and outputs of the individual algorithms that ALT offers which can be seen in figure 3.1. Apart from ALT algorithms it also includes finite automata editor done by Petr Svoboda [13].

This finite automata editor is called Statemaker and you can see a screenshot of how it looks in figure 3.2. To summarize its capabilities - users can:

- add states, as well as initial and final states
- add transitions between states
- edit transition string
- mark state as initial or final
- remove states and transitions
- import and export automaton in supported formats
- automatic positioning of transitions and states

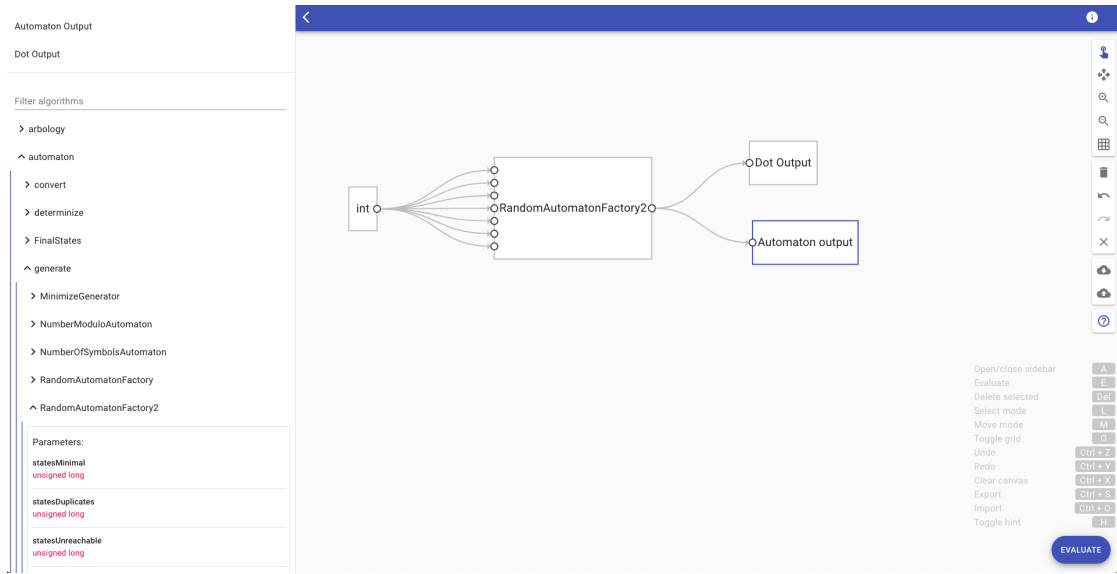


Figure 3.1 ALT web interface screenshot

All of the above features work reliably and are done in intuitive manner - user can quickly understand how to work with all the components. The most notable missing feature is easy simulation of input - this can be done via ALT web interface but if someone is looking for only editing FAs and simulating whether input string is accepted, they have to transition between two interfaces. The benefit is that they can then tap into all the other functionality that ALT offers. The author of Statemaker has chosen React and Typescript as underlying technologies [13].

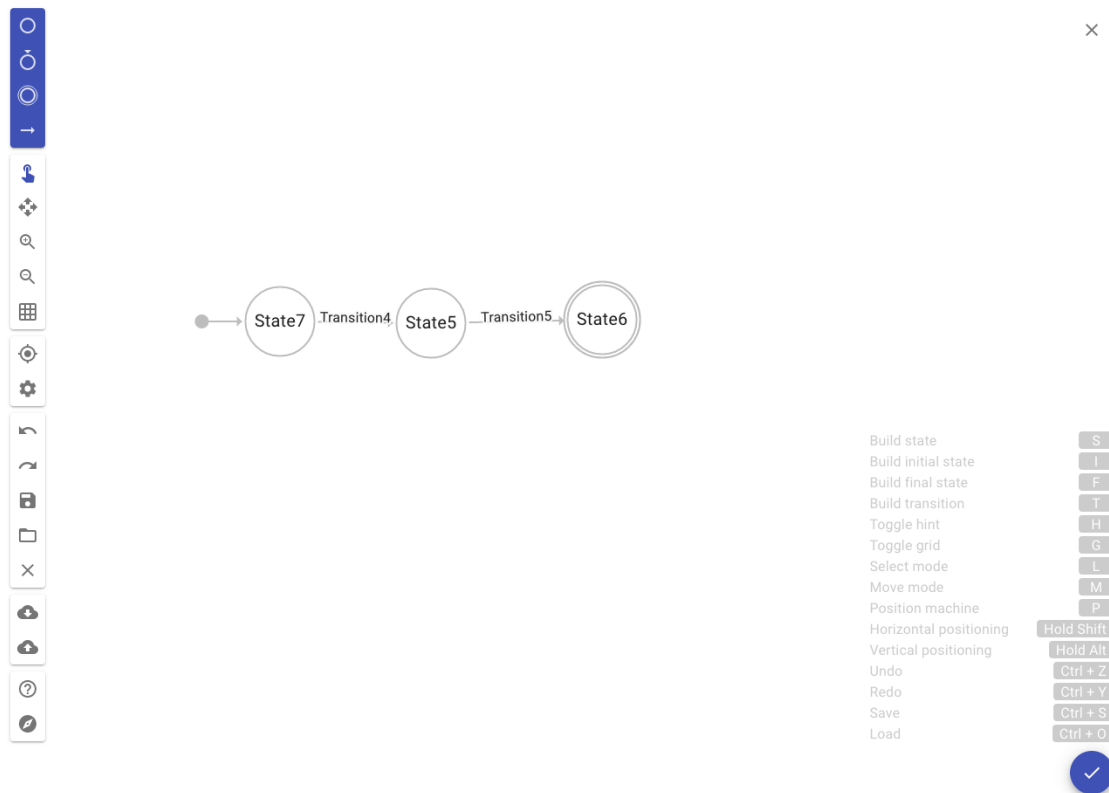
3.1.2 Existing Applications

As the main goal of this thesis is to write a finite automata editor for iPad, in this part I will study existing applications mainly for touch devices.

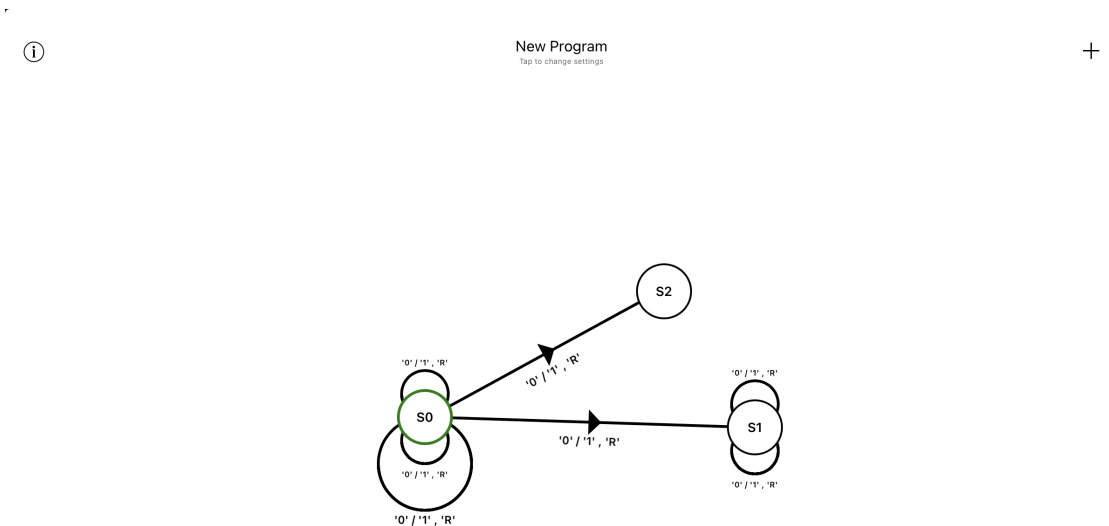
One of such applications is TuringSim [12]. Although, it is not for FAs but for a Turing machine, it also consists of an editor where user can add and edit states and transitions, thus making it similar to a FA editor. You can see its interface in figure 3.3. This editors lets users add and edit automaton's states and transitions. Users can also simulate input on the Turing machine's read-and-write tape. Editing of the automaton is done only via tap gestures which is similar to Statemaker with the difference that there are no distinct buttons for those actions. Therefore, it does not fully utilize the potential of touch devices as the UX is very similar to what would one experience on the web. Unfortunately, the app on iPad is broken at the moment as it is missing bottom toolbar for simulating input.

There is also app called Finite Automata [15]. In this app user can not edit automata in their weighted graph representation but instead has to use a command line that takes individual command which are described in the app. This app does not utilize touch device features at all.

There are also apps available as desktop applications. One is a Finite Automaton Editor by Jaime Rangel-Mondragon that is available as interactive Wolfram notebook [9]. This app allows you to edit the automaton via a transition table and does not allow to simulate any input. There is also Automata Editor by Max Shawabkeh [11]. In this desktop applications users can create and edit their automaton either via a table representation or regular expression. There are also features such as NFA determinization, evaluating automata on strings, and minimizing DFA. Thus it has a powerful feature set but one has to be already familiar with FA theory. It should also be noted that the feature set is a subset of what ALT web interface offers.



■ Figure 3.2 Statemaker screenshot



■ Figure 3.3 TuringSim interface screenshot

3.2 ALT Analysis

Now I will go over ALT and its features that can be leveraged for simulating FA input. The code of ALT is available on GitLab [4] where there are multiple repositories in group Algorithms Library Toolkit (webui-client, infrastructe, etc.) - that includes repository Algorithms Library Toolkit Core [1]. There we can find algorithms that can later be used for FA editor. The code is divided into multiple modules that are then built and linked together using CMake [7]. The most important module for this thesis are *alib2data* and *alib2algo* where *alib2data* contains FA models and *alib2algo* algorithms for simulating input.

3.3 FA Model

ALT offers

..... Chapter 4

Automata Editor Design

A decorative horizontal bar at the bottom of the page, composed of approximately 30 small blue squares arranged in a row.

Implementation



Chapter 6

User Testing

Chapter 7

Conclusion

7.1 Goals Assessment

Assess my goals.

7.2 Thesis Contribution

The contribution of this thesis is testing Algorithms Library Toolkit in practice and can now be pointed to for users who want to see its capabilities. The editor can now also be recommended in the course of BI-AAG at FIT CTU (and other universities) - for students and teachers alike.

7.3 Future Work

Algorithms Library Toolkit is an extensive library and there are still capabilities that are not implemented in the editor.

The choice of developing a native iOS application has resulted in good UX, but in the future it would be beneficial to broaden the possible audience and either develop a similar Android app or create a more ubiquitous web interface.

List of Abbreviations

DFA	Deterministic Finite Automaton
FA	Finite Automaton
NFA	Nondeterministic Finite Automaton
ALT	Algorithms Library Toolkit



Příloha A

Nějaká příloha

Sem přijde to, co nepatří do hlavní části.

Bibliography

- [1] *Algorithms Library Toolkit*. URL: <https://gitlab.fit.cvut.cz/algorithms-library-toolkit> (visited on 04/13/2020).
- [2] Patrick van Bergen. *Pipe-And-Filter*. URL: http://www.dossier-andreas.net/software_architecture/pipe_and_filter.html (visited on 04/10/2020).
- [3] Miroslav Čepěk and Daniel Vašata. *BIE- VZD*. URL: <https://courses.fit.cvut.cz/BIE-VZD/> (visited on 04/10/2020).
- [4] *GitLab*. URL: <https://about.gitlab.com/> (visited on 04/13/2020).
- [5] Jan Holub. *BIE-AAG*. URL: <https://courses.fit.cvut.cz/BIE-AAG/> (visited on 04/09/2020).
- [6] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Vol. 3. 2007. ISBN: 9780321455369.
- [7] Kitware. *CMake*. URL: <https://cmake.org/> (visited on 04/13/2020).
- [8] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. Dec. 2018. ISBN: 9780262039406.
- [9] Jaime Rangel-Mondragon. *A Finite Automaton Editor*. Mar. 2011. URL: <http://demonstrations.wolfram.com/AFiniteAutomatonEditor/> (visited on 04/13/2020).
- [10] Eliška Šestáková. *Automata and Grammars: A Collection of Exercises and Solutions*. Faculty of Information Technology, Czech Technical University in Prague, 2020. ISBN: 978-80-01-06462-7.
- [11] Max Shawabkeh. *Automata Editor*. URL: <http://max99x.com/school/automata-editor> (visited on 04/13/2020).
- [12] Cristian Stanescu. *TuringSim*. App available at <https://apps.apple.com/us/app/turingsim/id563350412>. URL: https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/ProjectDocuments/Progetto1_Stanesco.pdf (visited on 04/10/2020).
- [13] Petr Svoboda. *Webový editor konečných automat*. Bachelor's thesis. 2019.
- [14] Jan Trávníček, Tomáš Pecka, and Štěpán Plachý. *Algorithms Library Toolkit*. URL: <https://alt.fit.cvut.cz/> (visited on 04/06/2020).
- [15] Plinio Vilela. *Finite Automata*. URL: <https://apps.apple.com/us/app/finite-automata/id1043670880> (visited on 04/10/2020).
- [16] Michael Vrána. *Knihovna algoritmů ALT-webové rozhraní*. Bachelor's thesis. 2020.

Obsah přiloženého média

	readme.txt.....	stručný popis obsahu média
	exe.....	adresář se spustitelnou formou implementace
	src	
	impl.....	zdrojové kódy implementace
	thesis.....	zdrojová forma práce ve formátu L ^A T _E X
	text.....	text práce
	thesis.pdf.....	text práce ve formátu PDF