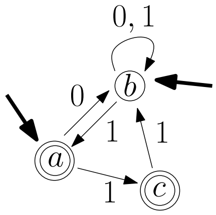
State Final Examination (Sample Questions)

Summer 2018

1 Automata (3 points)

1. Convert a nondeterministic finite state machine $A = (\{a, b, c\}, \{0, 1\}, \delta, \{a, b\}, \{a, c\})$ from the picture below to an equivalent deterministic finite state machine.
2. Classify the language $L(A)$ using the Chomsky hierarchy.



2 Algorithms and Data Structures: Text Search (3 points)

1. Define a searching automaton for the KMP (Knuth-Morris-Pratt) algorithm. Describe its states, forward and backward edges.
2. Draw a searching automaton for the string `KOCKODAN`.
3. What is the time complexity of the automaton construction algorithm and the time complexity of its application to find all occurrences of a word in a text?

3 Databases (3 points)

1. Sketch how you would represent an M:N binary relationship in a database table. Can you determine the original relationship cardinality (1:1, 1:N, M:N) when the relationship is represented by table $T(\underline{FK_1,FK_2})$? Explain.
2. Consider transactions T_1: $W(A)\ R(B)\ W(C)$ and T_2: $R(A)\ W(B)\ R(C)$. Is schedule S: $W_1(A)\ R_2(A)\ R_1(B)\ W_2(B)\ W_1(C)\ W_2(C)$ conflict serializable? Explain, and optionally propose an adjustment to make it so.

4 Object Oriented REST API Server Interface (3 points)

Consider a specialized web server for serving a REST API, implemented in a mainstream statically typed object oriented language (C++, C#, or Java) with the following features. The server supports dynamic installation of *services*. A service is an instance of a class implementing an `IService` interface, which encapsulates multiple REST *methods*. A service is identified by a string name, unique within the server scope. A REST method is also identified by a string name, unique within its service scope. The `IService` interface can provide a list of the supported REST methods. Exact relation between REST methods and HTTP methods (GET, POST) is not important. In our context, method is simply a member function of a class.

Request processing on the server starts with obtaining the URL and other request data from the HTTP protocol. The server parses this data to obtain the service identifier, the REST method identifier, and the list of arguments. Using the identifiers, the server looks up the requested service object and invokes the programming language method that implements the requested REST method. The programming language method receives the arguments in a collection such as a map or a dictionary (you can choose the type based on the selected programming language) where the keys are strings and values are objects of type `ParameterValue`, the return value is a string. Further details are not essential.

The server is represented by an object of class `Server`. Among other things, this class implements two essential methods with the following signature (the exact notation has to be specified based on the selected programming language):

```
public void add(IService service)
```

```
private string dispatch(string serviceId, string methodId, ParametersCollection parameters)
```

The `add` method registers a new service. This method can be relatively slow, since it is typically called solely during the server initialization (for example when the plugins are being loaded). The `dispatch` method is called internally during request processing to find and invoke the requested implementation method of the requested service. The `dispatch` method should have low overhead.

1. Propose the `IService` interface and briefly (for example using comments) explain the purpose of each method (unless it is quite clear from the signature).
2. Propose an implementation of the `IService` interface. Focus especially on the part of the interface responsible for reporting the supported REST methods.
3. Briefly (possibly in pseudocode) sketch the implementation of the `add` and `dispatch` methods of the `Server` class. If the methods access some member variables or fields, describe those too. Recall that the `dispatch` method should be reasonably efficient.

Minor syntax issues are not important, however, the overall design and logic should correspond to the features and the common sense use of your chosen programming language. If the language supports introspection or reflection, it can be used where appropriate.

5 File system (3 points)

This question contains a simplified FAT file system description. You can also consider a standard FAT file system in your answers (state explicitly if so).

Every partition formatted with the (simplified) FAT12/16/32 file system is divided into 3 parts:

1. Boot sector (logical sector number 0 of the partition),
2. So called FAT table, for definition see below (sectors 1 to N of the partition),
3. Data sectors (sectors N+1 to the last sector of the partition), these contain the actual data of the files stored in the file system (or data of directories – however, as directories do not differ from files in how their sectors are allocated in the file system, the rest of the assignment will not mention directories explicitly). For simplicity assume that the file system allocation unit is exactly one physical disk sector. Sectors have one of the commonly used sizes – let us mark this unit as X bytes.

Every data sector of a partition is assigned to exactly one file, or is marked as free. Every file in the file system occupies certain amount of whole sectors, however sectors of an individual file do not have to be allocated continuously (i.e. files can be fragmented). The number of a data sector (numbered from 1) containing the first X bytes of data of a file is stored in the directory entry describing the file (this is also the last file sector for files with sizes less than or equal to X). For files larger than X bytes the information about the following sectors is stored in the FAT table of the partition – for every file (with size of e.g. M sectors) the FAT table contains an “encoded” single linked list of a sequence of data sector numbers assigned to the file as its 2nd to Mth sector. The precise format of the FAT table is as follows:

For every data sector the FAT table contains exactly one Z-bit record (Z = 12 bits for FAT12, 16 bits for FAT16, 32 bits for FAT32) – this record is a Z-bit unsinged integer number stored in the little endian order. The FAT table contains no other information, so it is in fact an array of integer numbers. The record describing the data sector number 1 is located on offset 0 of the first FAT table sector, it is then immediately followed (without any padding) by the record for data sector 2, that is immediately followed by the record for data sector 3, and so on. If a record for data sector A contains value 0, then data sector A is free, and is not assigned to any file stored in the file system. If a record for data sector A contains value B, then the data sector A is assigned to some file, and the X bytes of this file’s data stored in sector A are followed by additional

X bytes of file's data stored in sector B (for fragmented files B does not have to be A+1, and in general B can be smaller or greater than A). If a record for data sector A contains the maximum value of a Z-bit number, then sector A is the last sector of a sequence of sectors assigned to some file (i.e. Mth sector of a file occupying M sectors in total). For example: if data of the A.TXT file is stored in data sectors 10, 7, 8, 15, then the directory entry for A.TXT will contain number 10 and the FAT table will contain the following: record for sector 10 will contain number 7, record for sector 7 will contain number 8, record for sector 8 will contain number 15, and record for sector 15 will contain a maximum value of a Z-bit number.

1. What can be a common value of X? Assume we have a partition with size of exactly 1 GiB. Should we choose the FAT16 or the FAT32 file system variant when formatting it? Explain why.
2. We stored 100 000 files, 1 KiB each, in the 1 GiB FAT-formatted partition. Now we want to read the contents of all these files. Should we prefetch the whole FAT table into memory, or should we read into memory only the relevant FAT sectors for every specific file just before reading the file's data? Explain why.
3. Assume we have a "global variable" `fat` (or a static field in a class) of the byte array type, into which the whole content of a FAT12-formatted partition's FAT table is loaded – one FAT table record has 12 bits, i.e. every 3 bytes of the `fat` array contain exactly 2 records describing 2 data sectors (the first sector number is in the first byte and the lower four bits of the second byte; the second sector number is in the upper four bits of the second byte and the third byte). Write a C#, Java, or C++ function with a single argument containing the number of the first data sector assigned to some file stored in a file system. The function will print the numbers of all the data sectors assigned to the same file (using the information stored in the `fat` variable).

6 HTTP Protocol (3 points)

This is a transcript of an HTTP session fragment:

```
GET / HTTP/1.1
Host: www.example.org
If-None-Match: "0123456789"
```

```
HTTP/1.1 200 OK
Date: Mon, 01 Jan 2000 00:00:00 GMT
Last-Modified: Mon, 01 Jan 2009 00:00:00 GMT
Cache-Control: max-age=3600
ETag: "9876543210"
Accept-Ranges: bytes
Content-Length: 123456
Content-Type: text/html

<!DOCTYPE html ...
```

1. What is the meaning of the first line of the request?
2. Why does the request contain the `Host` header?
3. What URL did the client likely use in the browser?
4. What other words can start the first line of the request and what is their meaning? At least two.
5. What is the meaning of the `Accept-Ranges` header in the response?
6. What is the meaning of the `If-None-Match` header in the request and how is it used?
7. What is the meaning of the `Last-Modified` header in the response and how is it used?

7 Optimization Methods (3 points)

Let $Ax \leq b$ be a linear inequality system in n variables. By multiplying each row by a positive constant we may assume that the first column of A is a vector with entries 0, -1 and 1 only. So we can write $Ax \leq b$ equivalently as

$$\begin{aligned} a_i'x' &\leq b_i \quad (i = 1, \ldots, m_1), \\ -x_1 + a_j'x' &\leq b_j \quad (j = m_1 + 1, \ldots, m_2), \\ x_1 + a_k'x' &\leq b_k \quad (k = m_2 + 1, \ldots, m), \end{aligned}$$

where $x' = (x_2, \ldots, x_n)$ and $a_1', \ldots, a_n'$ are the rows of A without the first entry. Then one can eliminate x_1: prove that $Ax \leq b$ has a solution if and only if the system

$$\begin{aligned} a_i'x' &\leq b_i & (i = 1, \ldots, m_1), \\ a_j'x' - b_j &\leq b_k - a_k'x' & (j = m_1 + 1, \ldots, m_2, k = m_2 + 1, \ldots, m) \end{aligned}$$

has a solution. Show that this technique, when iterated, leads to an algorithm for solving a linear inequality system $Ax \leq b$ (or proving infeasibility). What is the time complexity of the procedure?

8 Computational Linguistics: Formal Languages and Automata (specialization question – 3 points)

1. Describe non-projective dependency trees.
2. Give two examples of non-projective dependency trees.

9 Computational Linguistics: Basic Formalisms for Description of a Natural Language (specialization question – 3 points)

1. Describe basic facts about building language corpora.
2. Classify language corpora by annotation schemes. Give two examples for each category.
3. Explain what is valency in linguistics.

10 Computational Linguistics: Theory of Information (specialization question – 3 points)

We toss three fair (unbiased) coins. For each of them, heads and tails have the identical probability 1/2. The resulting random value is a vector $\langle x_1, x_2, x_3 \rangle$ where $x_i \in \{H, T\}$.

1. Compute entropy of this random variable.
2. How does the entropy change if one of the three coins is unfair (biased) and it always lands with a head up?

11 LA1 – Eigenvalues (3 points)

1. Let
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 3 & 0 & 1 \end{pmatrix}.$$
Decide whether $\lambda = 7$ is an eigenvalue of matrix A. Further decide whether $x = (1, 2, 1)^T$ is an eigenvector of matrix A.
2. Let V be the set of all 3×3 real matrices having eigenvalue 0 that corresponds to eigenvector $y = (1, 2, 3)^T$. Show that V is a vector space and determine its dimension.

12 LA2 – Positive definiteness (3 points)

Define the notion of positive definiteness of a matrix, and formulate at least one equivalent condition for checking positive definiteness of a matrix. Using this condition decide whether the following matrix is positive definite

$$B = \begin{pmatrix} 4 & -2 & 0 & 2 \\ -2 & 2 & 2 & -1 \\ 0 & 2 & 5 & -2 \\ 2 & -1 & -2 & 6 \end{pmatrix}$$

13 MA1 – Limit of a sequence (3 points)

Define the limit of a sequence.

Let $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ be sequences of reals. Decide whether it is true that if $\lim_{n\to\infty}(a_n - b_n) = 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Justify your decision.

14 MA2 – Metric space (3 points)

Let M be a non-empty set and f a function from M to positive reals. Show that for (M, d) is a metric space when d is a function defined as

$$d(x, y) = \begin{cases} 0 & x = y \\ \max(f(x), f(y)) & x \neq y. \end{cases}$$

15 MA3 – Tangent plane (3 points)

Find a point in which the tangent plane of the graph of a function $f(x, y) = x^2 + 2xy + 2y^2 - 6x - 8y$ is horizontal. What is an equation of the tangent plane ?

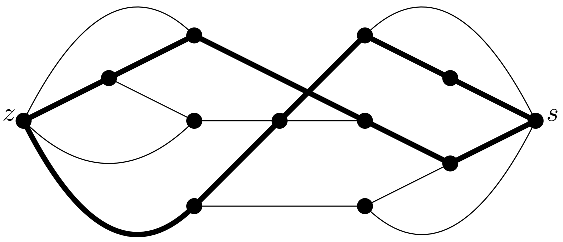
16 Independent random events (3 points)

1. Define when two random events are independent.
2. Let $n \geq 2$ be a positive integer. Construct a random graph with vertex set $\{1, 2, \ldots, n\}$ as follows: for each pair of vertices i and j such that $1 \leq i < j \leq n$ we toss a fair coin (which falls heads up with probability exactly 1/2) and if it falls heads up, we add the edge ij to G. Let J_i be the event "the degree of the vertex i is exactly 1". Determine for which n the two events J_1 and J_2 are independent. Justify your answer.

17 Graphs (3 points)

The graph G depicted below corresponds to a network, s.t. all edges are oriented from left to right and all capacities are 1.

A flow between z and s is indicated by bold edges, their capacity is fully used, while the other edges are not used at all.



Decide whether the flow is maximal. If it is not maximal, find a maximal flow and argue for its maximality.

What can be conluded from the value of the maximal flow about the connectivity of G ?

18 Combinatorics (3 points)

Let for $n \in \mathbb{N}$ the symbol $f(n)$ stand for the number of subsets of size 15 of an n-element set, and let $g(n)$ stand for the number of all mappings from an n-element set to a three-element set.

Express $f(n)$ and $g(n)$.

Decide, which of the following options holds:

- $\exists n_0 : \forall n > n_o : f(n) > g(n)$
- $\exists n_0 : \forall n > n_o : g(n) > f(n)$
- no n_0 suitable for at least one of the above two cases exists.

19 Logic (3 points)

Consider formulas φ_1, φ_2 in language $L = \langle 0, |\,|, f, -, < \rangle$ with equality, where 0 is a constant symbol, $|\,|, f$ are unary function symbols, $-$ is a binary function symbol and $<$ is a binary relation symbol

$$\begin{aligned} \varphi_1 : \quad & (\forall \varepsilon)(\forall u)(0 < \varepsilon \to (\exists \delta)(0 < \delta \wedge (\forall x)(|x-u| < \delta \to |f(x) - f(u)| < \varepsilon))), \\ \varphi_2 : \quad & (\exists u)(\exists \varepsilon)(0 < \varepsilon \wedge (\forall \delta)(0 < \delta \to (\exists x)(|x-u| < \delta \wedge \neg(|f(x) - f(u)| < \varepsilon)))). \end{aligned}$$

1. Give definitions for predicate logic when a formula φ is *valid in a structure* $\mathcal{A}$, when φ is *(logically) valid*, and when φ is *independent.*
2. Are formulas φ_1, φ_2 valid in a structure $\mathcal{A} = \langle \mathbb{R}, 0, |\,|, f, -, < \rangle$ of language L, where $0, |\,|, -, <$ has its usual meaning on $\mathbb{R}$ and $f(0) = 0$, $f(r) = |r|/r$ for $r \neq 0$? Give an explanation.
3. Is formula $\varphi_1 \leftrightarrow \neg\varphi_2$ independent ? Give an explanation.