Examination

Mathematical Foundations for Software Engineering

Course code DIT022

Date:	2018-10-30
Time:	14:00-18:00
Place:	Lindholmen
Teacher:	Christian Berger Michel Chaudron Richard Torkar
Visit to e*xam hall:*	14:30, 16:30
Questions:	6
Results:	Will be posted by 2018-11-20.
Grade Limits:	Pass (G) 50%, Pass with honors (VG) 90%
Allowed aids:	Calculators: Casio FX-82..., Texas TI-30... and Sharp EL-W531... Attached appendix with formulas and tables.

Please observe the following:

- <u>DO NOT</u> write your name on any answer sheet or exam sheet – write the anonymized code instead.
- Write in legible English (unreadable responses mean no points!).
- Motivate your answers and clearly state any assumptions made.
- Start each task on a new sheet!
- Write only on one side of the paper!
- Only answers written on the answer sheets will be graded, do not write on the exam sheets!
- Before handing in your exam, number and sort the sheets in task order!

NOTE:

Not following these instructions may result in the deduction of points!

Question 1 (1 + 1 + 2 + 4 + 1 + 14 = 23 pt)
"Automata and Grammars"

1.1 What language is generated by the given grammar G1? (1pt)
Note that we are not asking for partially correct solution(s) but for the fully correct one(s). Also, choosing multiple solutions, when only one is correct will not be assessed as correct. Mark clearly the correct solution.

G1 = (V, T, S, P) and V = {S, C, D, d, c}, where S is the start variable, T = {c, d} set of terminals and rules P:

S → Dc
D → Dd | Cd
C → Cc | c

a) $\mathcal{L}(G1) = \{c^m d^n c, where\ n, m \geq 1\}$
b) $\mathcal{L}(G1) = \{c^{m+1} d^n c, where\ n, m \geq 1\}$
c) $\mathcal{L}(G1) = \{c^m d^m c, where\ m \geq 0\}$
d) $\mathcal{L}(G1) = \{c^{\frac{m}{2}} d^n c, where\ n, m \geq 0\}$
e) $\mathcal{L}(G1) = \{c^{m+1} d^n c, where\ n \geq m, m > 0\}$

1.2 What language is generated by the given grammar G2? (1pt)
Note that we are not asking for partially correct solution(s) but for the fully correct one(s). Also, choosing multiple solutions, when only one is correct will not be assessed as correct. Mark clearly the correct solution.

G2 = (V, T, S, P) and V = {S, A, B, 0, 1}, where S is the start variable, T = {0, 1} set of terminals and rules P:

S → 0A | 1A
A → 0B
B → 1A | 1

a) $\mathcal{L}(G2) = \{11(01)^n, where\ n \geq 0\} \cup \{01(01)^m, where\ m \geq 1\}$
b) $\mathcal{L}(G2) = \{1(10)^m, where\ m \geq 1\} \cup \{0(01)^n, where\ n \geq 1\}$
c) $\mathcal{L}(G2) = \{00^m 1^n, where\ m, n \geq 0\} \cup \{(101)^n, where\ n \geq 1\}$
d) $\mathcal{L}(G2) = \{0(01)^n, where\ n \geq 1\} \cup \{1(01)^m, where\ m \geq 1\}$
e) $\mathcal{L}(G2) = \{001(0|1)^n, where\ n \geq 1\} \cup \{101(0|1)^m, where\ m \geq 1\}$

1.3 Provide two different non-empty words that are generated by the given grammar G5. (2pt)
*Note that **you should ignore** the need for spaces between the names and **assume** that the names need to be nonempty.*

G5 is given in the Backus-Naur form as:

```
<person> ::= <firstname><middleinitial><lastname>
<lastname> ::= <letterstring>
<middleinitial> ::= <letter>
<firstname> ::= <ucletter> | <ucletter><letterstring>
<letterstring> ::= <letter> | <letterstring><letter>
<letter> ::= <lcletter> | <ucletter>
<lcletter> ::= a | b | c |...| z
<ucletter> ::= A | B | C |...|Z
```

Notice that a production rule Nonterminal1 → Terminal | Nonterminal2 corresponds to a production rule `<Nonterminal1> ::= Terminal | <Nonterminal2>` in the Backus-Naur form.

1.4 Which production rules are missing in G6, so that it generates a language consisting of words described as: a lowercase letter, followed by a digit or an underscore, followed by three or four alphanumeric characters that could be lower or uppercase letters, or digits? (4pt)

G6:

```
<identifier> ::= <lcletter><digitorus><alphanumeric><alphanumeric><alphanumeric> |
<lcletter><digitorus><alphanumeric><alphanumeric><alphanumeric><alphanumeric>
<ucletter> ::= A | B | C |...|Z
<lcletter> ::= a | b | c |...|z
```

1.5 What does the deterministic finite automaton A1 do? (1pt)
Note that we are not asking for partially correct solution(s) but for the fully correct one(s). Also, choosing multiple solutions, when only one is correct will not be assessed as correct. Mark clearly the correct solution.

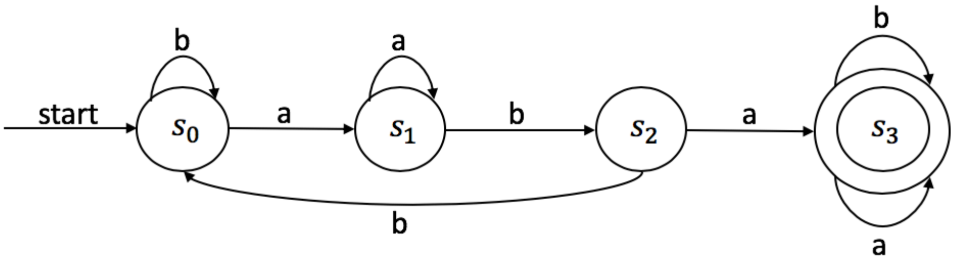

Figure 1: A1

a) $L(A1) = \{a, b\}^* aba\{a, b\}^*$
b) $L(A1) = \{b, a\}^+ ab^* a\{a, b\}^*$
c) $L(A1) = \{b, a\}^* aabba\{a, b\}^+$
d) $L(A1) = aba\{b, a\}^*$
e) $L(A1) = \{a, b\}^* ab^* a\{a, b\}^* b$

1.6 Draw a deterministic finite-state machine that recognizes the set of bit strings containing the set $\{\lambda, 0\} \cup \{0^m 1^n, where\ m \geq 1, n \geq 1\ and\ \lambda\ is\ an\ empty\ string\}$ (14 pt)
Note that unreadable drawings will be awarded with 0 points. Using wrong or incomplete notations will also result is deducing points!

Question 2 (8 + 1 + 1 + 7 = 17 pt)
"Logic"

2.1 Provide the complete truth table for the given compound preposition? (8pt)

$$(((v \to p) \lor t) \leftrightarrow p) \land (v \to (v \lor (t \land v)))$$

2.2 Express this statement using predicates and quantifiers. (1pt)
Note that we are not asking for partially correct solution(s) but for the fully correct one(s). Also, choosing multiple solutions, when only one is correct will not be assessed as correct. Mark clearly the correct solution.

Let $Q(x, y)$ be the statement "student x has been a contestant on quiz show y." Express S_1 in terms of $Q(x, y)$, quantifiers, and logical connectives, where the domain for x consists of all students at your school and for y consists of all quiz shows on television.

S_1 = "At least two students from your school have been contestants on Jeopardy."

a) $S_1 = \exists x_1 \exists x_2 (Q(x_1, Jeopardy) \wedge Q(x_2, Jeopardy))$
b) $S_1 = \exists x_1 \exists x_2 (Q(x_1 \wedge x_2, Jeopardy))$
c) $S_1 = \exists x_1 \exists x_2 (Q(x_1, Jeopardy) \wedge Q(x_2, Jeopardy) \wedge x_1 \neq x_2)$
d) $S_1 = \exists x_1 \exists x_2 \exists y_1 \exists y_2 (Q(x_1, y_1) \wedge Q(x_2, y_2) \wedge y_1 \neq y_2)$
e) $S_1 = \exists x_1 \exists x_2 \exists y_1 (Q(x_1, y_1) \wedge Q(x_2, y_1) \wedge x_1 \neq x_2)$

2.3 Inspect the system specification S_2 expressed with predicates and quantifiers. Precisely what does this specification mean in natural language? (1pt)

Note that we are not asking for partially correct solution(s) but for the fully correct one(s). Also, choosing multiple solutions, when only one is correct will not be assessed as correct. Mark clearly the correct solution.

$S_2 = \forall f\ (H(f) \rightarrow \exists c\ A(c))$, where $A(x)$ means that console x is accessible, and $H(x)$ means that fault condition x is happening.

a) S_2 = At least one console must be accessible during every fault condition.
b) S_2 = Several consoles must be accessible during every fault condition.
c) S_2 = During all fault conditions, all consoles must be accessible.
d) S_2 = There exists a fault condition for which none of the consoles are accessible.
e) S_2 = For all fault conditions one of the consoles is not accessible.

2.4 Look at the following compound proposition and answer the questions below. (7pt)

$$\big((P \land Q) \to \neg R\big) \leftrightarrow \big(P \lor (Q \to R)\big)$$

i) What is a tautology? (1pt)
ii) What is a contradiction? (1pt)
iii) Why is the given compound proposition not a tautology? (3pt)
iv) Why is the given proposition not a contradiction? (1pt)
v) Turn the given compound proposition into a tautology by making changes to the compound proposition (for example, changing an $\land$ to $\lor$, but not removing individual propositions p, q and r). (1pt)

Question 3 (5 + 5 = 10 pt)

"Proofs"

3.1 Let $n \in \mathbb{N}, n > 1$. If we suppose that n is not prime then $2^n - 1$ is not a prime. Prove this statement by using the direct proof method. (5pt)

Notice that:

- Since n is not a prime, $\exists a, b \in \mathbb{N}$ such that $n = a * b,\ where\ 1 < a, b < n$.
- $t^m - 1 = (t-1)(1 + t + t^2 + \cdots + t^{m-1})$ thus, if $t = 2^b$ and $m = a$ then $2^{ab} - 1 = (2^b - 1)(1 + 2^b + 2^{2b} + \cdots + 2^{(a-1)b})$.

3.2 Prove $1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}$ by using the mathematical induction method. (5pt)

Question 4 (1 + 5 + 3 + 6 = 15 pt)
"Program Complexity"

4.1 Determine the complexity of the following code snippet in terms of Big-O. (1pt)

```
int i = 0;
while (i < n) {
        found = false;
        j = 0;
        while (j < n) {
                if (a[i] == b[j]) {
                        found = true;
                        break; // break means leaving the current loop
                }
                ++j;
        }
        if (found == false) {
                return false;
        }
        ++i;
}
return true;
```

4.2 Assuming that the function call *print* is of *O(1)* complexity, determine the complexity of methods 1,2,3, and 4 in terms of Big-O. How many times will *method3* print the array? (5pt)

```
void method1(int[] arr) {
        int n = arr.length;
        for (int i = n - 1; i >= 0; i = i - 3) {
                print(arr[i]);
        }
}


void method2(int[] arr) {
        for (int i = 0; i < arr.length(); i++) {
                for (int k = 0; k < arr.length(); k++) {
                        print(Math.log(arr[i]));
                }
        }
}


void method3(int[] arr) {
        for (int i = 0; i < arr.length(); i++) {
                method1(arr);
                method2(arr);
        }
}


void method4(int[] arr) {
        for (int i = 0; i < arr.length(); i++) {
                for (int k = arr.length() - 1; k > 0; k = k/3) {
                        print(arr[i]);
                }
        }
}
```

4.3 Calculate the Big-O complexity of each of these functions. (3pt)

i. $(n^3 + n^2 logn)(logn + 1) + (17logn + 19)(n^3 + 2)$
ii. $(2n + n^2)(n^3 + 3^n)$
iii. $(n^n + n2^n + 5^n)(n! + 5^n)$

4.4 Look at the array below and consider how the array changes at each step of going through a selection-sort program, where the values of the array are being sorted in an increasing order (i.e. 1 2 3). Write down the state of the array BEFORE and AFTER for each loop iteration of the program. What does the array look like after 3 iterations of selection-sort? (6pt)

Note that "after 3 iterations" means, when i=2, before entering the for loop and increasing i to 3. Assume that i begins with 0.

Array: 200, 67, 4, 192, 12, 94, 1, 88, 9, 43

Question 5 (3 + 1 + 4 + 2 = 10 pt)
"Graph and Set Theory"

5.1 Draw an undirected graph represented by the give adjacency matrix A in Figure 8. (3pt)

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 0 & 4 \\ 2 & 4 & 0 \end{bmatrix}$$

Figure 2: The adjacency matrix of an undirected graph.

5.2 Look at the graphs G_1 with a set of vertices $\{u_1, u_2, \dots, u_{10}\}$ and G_2 with a set of vertices $\{v_1, v_2, \dots, v_{10}\}$ as shown in Figure 9. Are these graphs isomorphic? Provide an argument for your claim. (1pt)

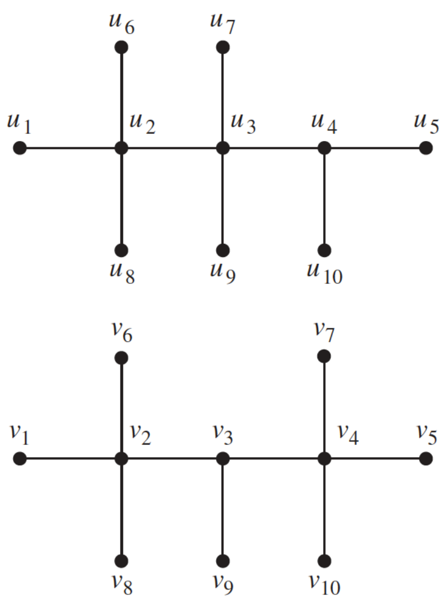

Figure 9: Graphs G_1 (top) and G_2 (bottom).

5.3 Look at the graph G_3 depicted in Figure 10 and answer the following questions. (4pt)

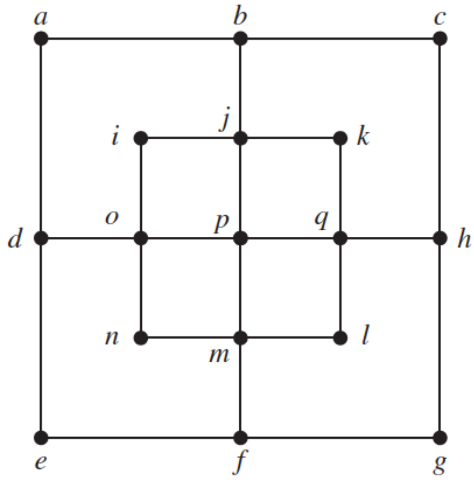

Figure 10: Graph G_3.

i. Is the graph connected? Provide an argument for your claim. (1pt)
ii. What are the maximum and minimum degrees of vertices from G_3? (1pt)
iii. Does G_3 have a Hamilton circuit? If it does, find the circuit and write it down. If it does not, give an argument to show why such a circuit cannot exist. (1pt)
iv. Does G_3 have a Hamilton path? If it does, find the path and write it down. If it does not, give an argument to show why such a path cannot exist. (1pt)

5.4 Look at the graph G_4 depicted in Figure 11. Does this graph have an Euler path? If it does, find the path and write it down. If it does not, provide an argument for your claim. (2pt)

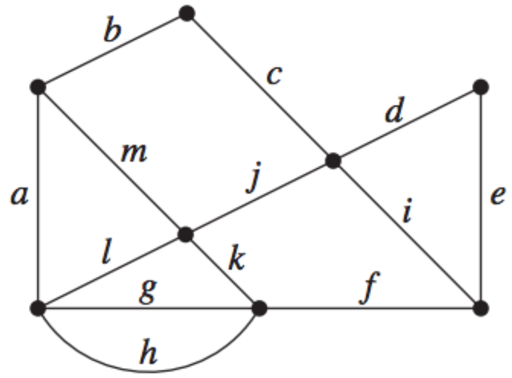

Figure 10: Graph G_4.

Question 6 (3 + 3 + 7 + 9 + 3 = 25 pt)

"Statistics"

6.1 A group of 4 girls and 16 boys is lined up in random order - that is, each of the 20! permutations is assumed to be equally likely. (3pt)

i. What is the probability that the person in the 2nd position is a boy? (Write answer in fractional form, e.g., 1/2)
ii. What about the person being a girl in the 12th position? (Write answer in fractional form, e.g., 1/2)
iii. Finally, what is the probability that a particular boy is in the 3rd position? (Write answer in fractional form, e.g., 1/2)

6.2 Richard is a randomly chosen member of a large population in which we assume that there are 1% terrorists. The police arrest Richard and the court rules that Richard is a terrorist and needs to serve time in prison. The court correctly sentences terrorist in 95% of the cases and correctly identifies non-terrorists 90% of the time, i.e., releases them. (3pt)

To determine the probability that Richard is a terrorist (T) **and** the court finds this to be true (C), we can apply Bayes' theorem, which in its general form states:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What is the probability that Richard is a terrorist and that the court finds him guilty?

6.3 Perform **calculations and find** the linear regression $Y = \alpha + \beta x$ for the data in the table below. **Account** for how you deal with missing y values! **Explain** what α, β mean if you use them to plot a line! (7pt)

X	Y
645	3.0
646	3.4
647	3.5
648	3.9
649	-
650	4.1
651	-
652	4.6

- no data registered.

6.4 Look at the collected data from three different groups below and answer the following questions about the differences between these groups. (9pt)

X_1: 3, 5, 8
X_2: 3, 5, 6
X_3: 2, 6, 6

i. Calculate the total sum of squares (SST), within sum of squares (SSW), and between sum of squares (SSB) and conduct inferential statistics. (3pt)
ii. What are the degrees of freedom for SST, SSW, and SSB? (3pt)
iii. What is the null hypothesis? (1pt)
iv. Calculate the F-statistic. Do we have a statistically significant difference on the 0.1 level if the critical level is 3.46 (we do not talk about the p-value now)? (2pt)

6.5 Select the correct answers (can be more than one!). You get 0 points if you pick a wrong answer. (3pt)

i. Data often contains outliers. One can deal with them in different ways. How will you deal with outliers the coming three years?
   a) Remove them.
   b) Use unsupervised outlier detection in order to get an unbiased result.
   c) Never remove them.
ii. We want to determine if a sample is normally distributed, hence we can use:
   a) Normal probability plots.
   b) Kolmogorov-Smirnov test for normality.
   c) Nothing.
iii. We listened carefully during Richard's lecture and we now understand that we need to determine the sample size of our study design:
   a) A priori.
   b) A posteriori.
   c) During data collection.