HW1CSE105W25: Homework assignment 1

CSE105W25

Due: January 16th at 5pm, via Gradescope

In this assignment,

You will practice reading and applying the definitions of alphabets, strings, languages, Kleene star, and regular expressions. You will use regular expressions and relate them to languages.

Resources: To review the topics for this assignment, see the class material from Weeks 0 and 1 and Review Quiz 1. We will post frequently asked questions and our answers to them in a pinned Piazza post.

Reading and extra practice problems: Sipser Section 0, 1.3. Chapter 0 exercises 0.1, 0.2, 0.3, 0.5, 0.6, 0.9. Chapter 1 exercises 1.19, 1.23.

For all HW assignments: Weekly homework may be done individually or in groups of up to 3 students. You may switch HW partners for different HW assignments. Please ensure your name(s) and PID(s) are clearly visible on the first page of your homework submission and then upload the PDF to Gradescope. If working in a group, submit only one submission per group: one partner uploads the submission through their Gradescope account and then adds the other group member(s) to the Gradescope submission by selecting their name(s) in the "Add Group Members" dialog box. You will need to re-add your group member(s) every time you resubmit a new version of your assignment. Each homework question will be graded either for correctness (including clear and precise explanations and justifications of all answers) or fair effort completeness. On the "graded for correctness" questions, you may only collaborate with CSE 105 students in your group; if your group has questions about a problem, you may ask in drop-in help hours or post a private post (visible only to the Instructors) on Piazza. On the "graded for completeness" questions, you may collaborate with all other CSE 105 students this quarter, and you may make public posts about these questions on Piazza.

All submitted homework for this class must be typed. You can use a word processing editor if you like (Microsoft Word, Open Office, Notepad, Vim, Google Docs, etc.) but you might find it useful to take this opportunity to learn LaTeX. LaTeX is a markup language used widely in computer science and mathematics. The homework assignments are typed using LaTeX and you can use the source files as templates for typesetting your solutions. To generate state diagrams of

machines, you can (1) use the LaTex tikzpicture environment (see templates in the class notes), or (2) use the software tools Flap.js or JFLAP described in the class syllabus (and include a screenshot in your PDF), or (3) you can carefully and clearly hand-draw the diagram and take a picture and include it in your PDF. We recommend that you submit early drafts to Gradescope so that in case of any technical difficulties, at least some of your work is present. You may update your submission as many times as you'd like up to the deadline.

Integrity reminders

- Problems should be solved together, not divided up between the partners. The homework is designed to give you practice with the main concepts and techniques of the course, while getting to know and learn from your classmates.
- On the "graded for correctness" questions, you may only collaborate with CSE 105 students in your group. You may ask questions about the homework in office hours (of the instructor, TAs, and/or tutors) and on Piazza (as private notes viewable only to the Instructors). You *cannot* use any online resources about the course content other than the class material from this quarter – this is primarily to ensure that we all use consistent notation and definitions (aligned with the textbook) and also to protect the learning experience you will have when the 'aha' moments of solving the problem authentically happen.
- Do not share written solutions or partial solutions for homework with other students in the class who are not in your group. Doing so would dilute their learning experience and detract from their success in the class.

You will submit this assignment via Gradescope (https://www.gradescope.com) in the assignment called "hw1CSE105W25".

Assigned questions

1. **Strings and languages: finding examples and edge cases** (12 points):

(a) (*Graded for completeness*) [1] Give five (different) example alphabets that are meaningful or useful to you in some way. Specify them formally, either with roster notation (which means listing all and only distict elements between { and } and separated by commas) or with another approach to precisely define all and only the elements of the alphabet.

(b) (*Graded for completeness*) Give an example of a finite set over an alphabet and an infinite set over an alphabet. You get to choose the alphabet, and you get to choose the sets.

[1] This means you will get full credit so long as your submission demonstrates honest effort to answer the question. You will not be penalized for incorrect answers. To demonstrate your honest effort in answering the question, we expect you to include your attempt to answer *each* part of the question. If you get stuck with your attempt, you can still demonstrate your effort by explaining where you got stuck and what you did to try to get unstuck.

The goal is to practice communicating your choices and definitions with clear and precise notation. One habit that will be useful (for this course, and beyond), is to think of your response for each question as a well-formed paragraph: include all the information that is relevant so that your solution is self-contained, and so that each sentence is grammatically constructed.

(c) (*Graded for correctness*) [2] Define an alphabet Σ_1 and an alphabet Σ_2 and a language L_1 over Σ_1 that is also a language over Σ_2 and a language L_2 over Σ_2 that is **not** a language over Σ_1. A complete and correct answer will use clear and precise notation (consistent with the textbook and class notes) and will include a description of why the given example L_1 is a language over both Σ_1 and Σ_2 and a description of why the given example L_2 is a language over Σ_2 and not over Σ_1.

2. **Regular expressions** (20 points):

(a) (*Graded for completeness*) Give three regular expressions that all describe the set of all strings over $\{a, b\}$ that have odd length. Ungraded bonus challenge: Make the expressions as different as possible!

(b) (*Graded for completeness*) A friend tells you that each regular expression that has a Kleene star (*) describes an infinite language. Are they right? Either help them justify their claim or give a counterexample to disprove it and explain your counterexample.

(c) (*Graded for correctness*) For this question, the alphabet is $\{a, b, c\}$. A friend is trying to design a regular expression that describes the set of all strings over this alphabet that end in c. Classify each of the following attempts as

- Correct. Explain why.
- Error Type 1: Incorrect, because (even though each string that is in the language described by the regular expression ends in c) there is a string that ends in c that is not in the language described by the regular expression. Give this example string and explain why it proves we're in this case.
- Error Type 2: Incorrect, because (even though each string that ends in c is in the language described by the regular expression), there is a string in the language described by the regular expression that does not end in c. Give this example string and explain why it proves we're in this case.
- Error Type 3: Incorrect, because there are two counterexample strings, one which is a string that ends in c that is not in the language described by the regular expression and one which is in the language described by the regular expression but does not end in c.Give both example strings and describe why each has the given property.

[2] This means your solution will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should explain how you arrived at your conclusions, using mathematically sound reasoning. Whether you use formal proof techniques or write a more informal argument for why something is true, your answers should always be well-supported. Your goal should be to convince the reader that your results and methods are sound.

Worked example for reference: Consider the regular expression $(a \cup b \cup c)^*$. This regular expression has **Error Type 2** because it describes the set of all strings over $\{a, b, c\}$, so even though each string that ends in c is in this language, there is an example, say ab that is a string in the language described by the regular expression (because we consider the string formed as a result of the Kleene star operation which has 2 slots and where the first slot matches the a in $a \cup b \cup c$ and the second slot matches b in $a \cup b \cup c$) but does not end in c (it ends in b).

i. The regular expression is

$$(a \cup b)^* \circ c$$

ii. The regular expression is

$$(a \circ b \circ c)^*$$

iii. The regular expression is

$$a^*c \ \cup \ b^*c \ \cup \ c^*c$$

3. **Functions over languages** (18 points):

For each language L over an alphabet Σ, we have the associated sets of strings (also over Σ)

$$L^* = \{w_1 \cdots w_k \mid k \geq 0 \text{ and each } w_i \in L\}$$

and

$$SUBSTRING(L) = \{w \in \Sigma^* \mid \text{there exist } x, y \in \Sigma^* \text{ such that } xwy \in L\}$$

and

$$EXTEND(L) = \{w \in \Sigma^* \mid w = uv \text{ for some strings } u \in L \text{ and } v \in \Sigma^*\}$$

Also, recall the set operations union and intersection: for any sets X and Y

$$X \cup Y = \{w \mid w \in X \text{ or } w \in Y\}$$

$$X \cap Y = \{w \mid w \in X \text{ and } w \in Y\}$$

(a) (*Graded for completeness*) Specify an example language A over $\{0, 1\}$ such that

$$SUBSTRING(A) = EXTEND(A)$$

or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language A and a precise and clear description of the result of computing $SUBSTRING(A)$, $EXTEND(A)$ (using the given definitions) to justify this description and to justify the set equality, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

(b) (*Graded for correctness*) Specify an example language B over $\{0, 1\}$ such that

$$SUBSTRING(B) \cap EXTEND(B) = \{\varepsilon\}$$

and

$$SUBSTRING(B) \cup EXTEND(B) = \{0,1\}^*$$

or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language B and a precise and clear description of the result of computing $SUBSTRING(B)$, $EXTEND(B)$ (using the given definitions) to justify this description and to justify the set equality with $\{\varepsilon\}$ and $\{0,1\}^*$ (respectively), or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

(c) (*Graded for correctness*) Specify an example **infinite** language C over $\{0,1\}$ such that

$$SUBSTRING(C) \neq \{0,1\}^*$$

and

$$SUBSTRING(C) = C^*$$

or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language C and a precise and clear description of the result of computing $SUBSTRING(C)$, C^* (using the given definitions) to justify this description and to justify the set nonequality claims, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.

(d) (*Graded for correctness*) Specify an example **finite** language D over $\{0,1\}$ such that

$$EXTEND(D) \neq \{0,1\}^*$$

and

$$EXTEND(D) = D^*$$

or explain why there is no such example. A complete solution will include either (1) a precise and clear description of your example language D and a precise and clear description of the result of computing $EXTEND(D)$, D^* (using the given definitions) to justify this description and to justify the set nonequality claims, or (2) a sufficiently general and correct argument why there is no such example, referring back to the relevant definitions.