Section 2: Grammars & Ambiguity

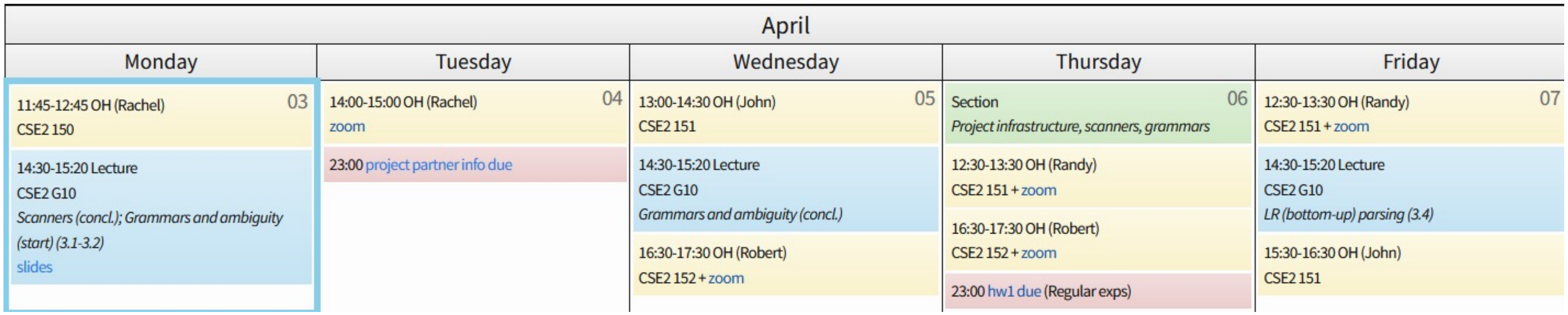
CSE 401/M501

Adapted from Spring 2021

Announcements

- Due ***Tonight*** at **11PM**: HW1
- Due Thursday 4/13 at 11PM: scanner part of project
  - You'll be using git/CSE GitLab for project
  - Remember to **git tag** your submission
- Office hours on Zoom (Canvas tab) and in person

WE ARE HERE



April

Monday	Tuesday	Wednesday	Thursday	Friday
03	04	05	06	07
11:45-12:45 OH (Rachel) CSE2 150	14:00-15:00 OH (Rachel) zoom	13:00-14:30 OH (John) CSE2 151	Section *Project infrastructure, scanners, grammars*	12:30-13:30 OH (Randy) CSE2 151 + zoom
14:30-15:20 Lecture CSE2 G10 *Scanners (concl.); Grammars and ambiguity (start) (3.1-3.2)* slides	23:00 project partner info due	14:30-15:20 Lecture CSE2 G10 *Grammars and ambiguity (concl.)*	12:30-13:30 OH (Randy) CSE2 151 + zoom	14:30-15:20 Lecture CSE2 G10 *LR (bottom-up) parsing (3.4)*
		16:30-17:30 OH (Robert) CSE2 152 + zoom	16:30-17:30 OH (Robert) CSE2 152 + zoom	15:30-16:30 OH (John) CSE2 151
			23:00 hw1 due (Regular exps)	

Agenda

- Git Review
- Walkthrough of starter code
- Grammar/Ambiguity Practice

Git Review

https://drive.google.com/file/d/1KdoFaMC8ppjTepgpk7QNzkUrZtmH32WZ/view

Git Review – SSH Keys

- An SSH key lets a `git` server remember a specific client computer
- If `git` asks for a password to push or pull, you need to setup an SSH key
- Typically just need to do the following:
  - `ssh-keygen -t rsa -C "you@cs.washington.edu" -b 4096`
  - Copy `~/.ssh/id_rsa.pub` into your GitLab account
- Full setup and troubleshooting instructions: https://gitlab.cs.washington.edu/help/user/ssh.md

Git Review – Version Control

- The “official” repo (a.k.a., the **remote**) lives on the CSE GitLab server
- **Cloning** a repo gives you a private, local *copy*
- **Committing** saves *local* changes into the *local* repo’s revision history
- **Push** to send *local* commits to *remote* repo
- **Pull** to bring *remote* commits to *local* repo
- Beware of **merge conflicts** – pull frequently

Git Review – Version Control

- When in doubt, Google it and consider the StackOverflow answer with the most upvotes
  - Make sure to check the replies for words of caution
  - Be *absolutely sure* it applies in our situation
  - Be careful of using the 'force' option; don't rebase; don't branch; don't get cute or clever
  - Don't hesitate to copy-paste a backup somewhere before messing around with weird git commands
- We're here to help
- Consult the official git documentation at https://git-scm.com/doc

Git Review – The 401 Repository

- Each project pair is given a repository in which to work and collaborate
  - The repository starts out with a tiny demo compiler to show how the tools work together
- You will submit each phase of the project using a tag in the repository (see each project phase spec for exact tag)
- To get started, simply clone your repo locally and get started!

Code Walkthrough!

https://drive.google.com/file/d/1ZL2h5Pn96nPbEJ-LCczZMQxbQT56LmH2/view

Summary: Project Structure

- Use ant to clean/compile/test…
- See README.txt for full folder description
  - src: your minijava compiler code
    - DemoParser.java and DemoScanner.java: example usages for you
    - MiniJava.java: the main compiler file, you will create this file and build on it for each lab
    - Scanner/minijava.jflex: Scanner code (bulk of lab 1)
    - Parser/minijava.cup: Parser code (bulk of lab 2)
    - Note: don’t push build files; run `ant clean`
  - test: tests you will write
    - junit: JUnit tests for minijava
    - resources: your minijava programs and expected output
  - SamplePrograms: example programs for you

Summary: to support a new token

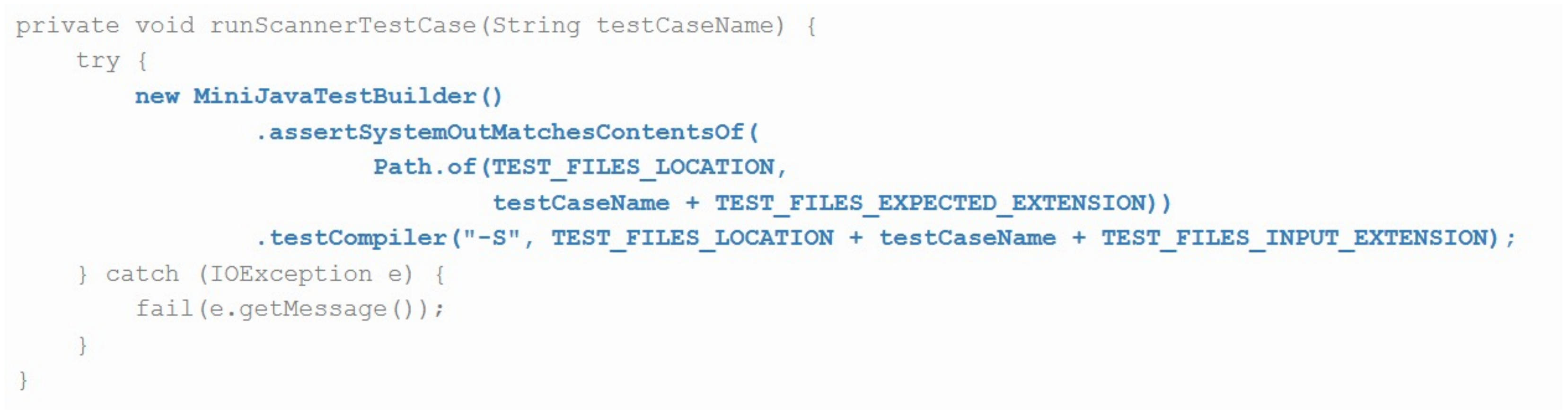
- src/Parser/minijava.cup
  - Add a new terminal for the symbol
- src/Scanner/minijava.jflex
  - Add a new regex rule to return the new symbol on match
  - If you want the raw value
    - Add a new case in `symbolToString`
    - Use `yytext()` to get the raw value

To avoid the common mistakes…

- Implement MiniJava, break the demo code/tests if needed
  - Read input from the specified file (**NOT** `System.in`), print output to `System.out`
  - Print errors to `System.err`
  - Use `System.exit` with status 1 after processing **entire file** if errors; status 0 if none
- Write and run (a lot of) junit tests
  - … and double check with the MiniJava grammar
- Do **NOT** modify or commit the generated files
  - Run `ant clean` before commit

Optional Testing Framework

- Experimental framework new to 22au (credit: Apollo Zhu)
- Simplifies the test code for MiniJava:



```
private void runScannerTestCase(String testCaseName) {
    try {
        new MiniJavaTestBuilder()
                .assertSystemOutMatchesContentsOf(
                        Path.of(TEST_FILES_LOCATION,
                                testCaseName + TEST_FILES_EXPECTED_EXTENSION))
                .testCompiler("-S", TEST_FILES_LOCATION + testCaseName + TEST_FILES_INPUT_EXTENSION);
    } catch (IOException e) {
        fail(e.getMessage());
    }
}
```

- Allows for testing error output and exit codes too
- Check out the website for more details on how to use this tool!

Grammar Worksheet!

Answers

Problem 1a

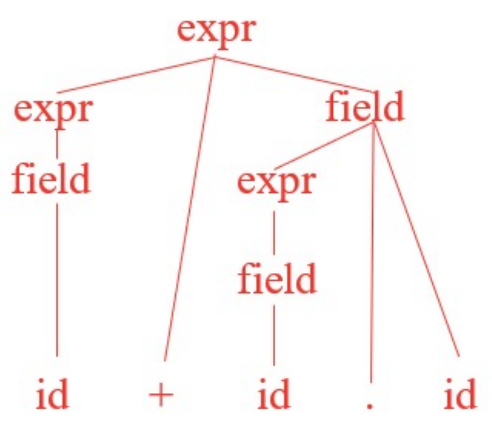
1) Consider the following syntax for expressions involving addition and field selection:

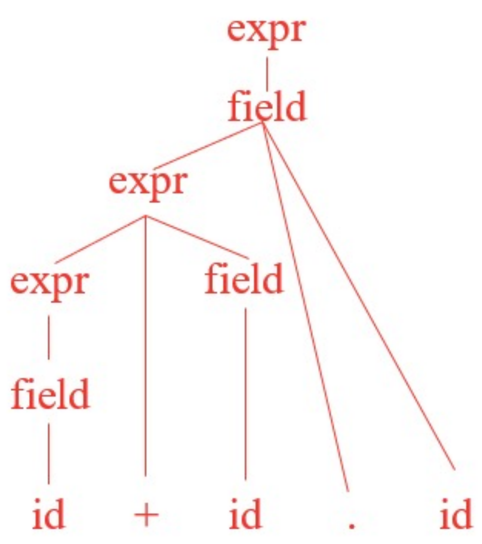
$$expr ::= expr + field$$
$$expr ::= field$$
$$field ::= expr\ .\ \mathrm{id}$$
$$field ::= \mathrm{id}$$

a) Show that this grammar is ambiguous.

Problem 1a solution

Here are two derivations of id+id.id:





Problem 1b

1b) Give an unambiguous context-free grammar that fixes the problem(s) with the grammar in part (a) and generates expressions with `id`, field selection, and addition. As in Java, field selection should have higher precedence than addition and both field selection and addition should be left-associative (i.e. `a+b+c` means `(a+b)+c`).

expr ::= *expr* + *field*

expr ::= *field*

field ::= *expr* . id

field ::= id

Problem 1b answer

1b) Give an unambiguous context-free grammar that fixes the problem(s) with the grammar in part (a) and generates expressions with `id`, field selection, and addition. As in Java, field selection should have higher precedence than addition and both field selection and addition should be left-associative (i.e. `a+b+c` means `(a+b)+c`).

The problem is in the first rule for *field*, which creates an ambiguous precedence

expr ::= *expr* + *field*

expr ::= *field*

field ::= ***field*** . id

field ::= id

Problem 2

2) The following grammar is ambiguous:

A ::= B b C

B ::= b | ε

C :: = b | ε

To demonstrate this ambiguity we can use pairs of derivations. Here are five different pairs. For each pair of derivations, circle OK if the pair correctly proves that the grammar is ambiguous. Circle WRONG if the pair does *not* give a correct proof. You do not need to explain your answers.

(Note: Whitespace in the grammar rules and derivations is used only for clarity. It is not part of the grammar or of the language generated by it.)

Problem 2a

$A ::= B \text{ b } C$
$B ::= \text{b} \mid \varepsilon$
$C ::= \text{b} \mid \varepsilon$

2a)

$\mathrm{A} \Rightarrow B \text{ b } C \Rightarrow \text{b b } C \Rightarrow \text{b b b}$
$\mathrm{A} \Rightarrow B \text{ b } C \Rightarrow B \text{ b b} \Rightarrow \text{b b b}$

Problem 2a answer

$A ::= B \; b \; C$

$B ::= b \mid \varepsilon$

$C ::= b \mid \varepsilon$

2a)

A => B b C => b b C => b b b

A => B b C => B b b => b b b

Wrong: Mix of left/rightmost derivations; also b b b has unique leftmost and unique rightmost derivations

Problem 2b

2b)

$A \Rightarrow B\ b\ C \Rightarrow b\ b\ C \Rightarrow b\ b$

$A \Rightarrow B\ b\ C \Rightarrow b\ C \Rightarrow b\ b$

$A ::= B\ b\ C$

$B ::= b \mid \varepsilon$

$C ::= b \mid \varepsilon$

Problem 2b answer

$A ::= B \text{ b } C$
$B ::= \text{b} \mid \varepsilon$
$C ::= \text{b} \mid \varepsilon$

2b)

A => B b C => b b C => b b
A => B b C => b C => b b

Ok: Two different leftmost derivations of b b

Problem 2c

$A ::= B \text{ b } C$
$B ::= \text{b} \mid \varepsilon$
$C ::= \text{b} \mid \varepsilon$

2c)

$\text{A} \Rightarrow B \text{ b } C \Rightarrow \text{b b } C \Rightarrow \text{b b}$

$\text{A} \Rightarrow B \text{ b } C \Rightarrow B \text{ b b} \Rightarrow \text{b b}$

Problem 2c answer

$A ::= B\ \mathrm{b}\ C$
$B ::= \mathrm{b} \mid \varepsilon$
$C ::= \mathrm{b} \mid \varepsilon$

2c)

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}$

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow B\ \mathrm{b}\ \mathrm{b} \Rightarrow \mathrm{b}\ \mathrm{b}$

Wrong: Different derivations: one leftmost, one rightmost

Problem 2d

$A ::= B\ \mathrm{b}\ C$

$B ::= \mathrm{b} \mid \varepsilon$

$C ::= \mathrm{b} \mid \varepsilon$

2d)

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}$

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ \mathrm{b}$

Problem 2d answer

$A ::= B\ \mathrm{b}\ C$
$B ::= \mathrm{b} \mid \varepsilon$
$C ::= \mathrm{b} \mid \varepsilon$

2d)

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}$

$\mathrm{A} \Rightarrow B\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ C \Rightarrow \mathrm{b}\ \mathrm{b}\ \mathrm{b}$

Wrong: Two different strings, not two derivations of same string

Problem 2e

2e)

$A \Rightarrow B\ b\ C \Rightarrow B\ b \Rightarrow b\ b$

$A \Rightarrow B\ b\ C \Rightarrow B\ b\ b \Rightarrow b\ b$

$A ::= B\ b\ C$

$B ::= b \mid \varepsilon$

$C ::= b \mid \varepsilon$

Problem 2e answer

2e)

$A \Rightarrow B\ b\ C \Rightarrow B\ b \Rightarrow b\ b$

$A \Rightarrow B\ b\ C \Rightarrow B\ b\ b \Rightarrow b\ b$

$A ::= B\ b\ C$

$B ::= b \mid \varepsilon$

$C ::= b \mid \varepsilon$

Ok: Two different rightmost derivations of b b

Problem 3

3) The following grammar is ambiguous. (As before, whitespace is used only for clarity; it is not part of the grammar or the language generated by it.)

P ::= ! Q | Q && Q | Q
Q ::= P | id

Give a grammar that generates exactly the same language as the one generated by this grammar but that is not ambiguous. You may resolve the ambiguities however you want – there is no requirement for any particular operator precedence or associativity in the resulting grammar.

Problem 3 answer

3) Original grammar:

$$P \ ::= \ !\,Q \mid Q \,\&\&\, Q \mid Q$$
$$Q \ ::= \ P \mid \text{id}$$

This solution disambiguates ! and && by putting them in different productions, and also forces the binary operator && to be left-associative:

$$P \ ::= \ P \,\&\&\, Q \mid Q$$
$$Q \ ::= \ !Q \mid \text{id}$$

Other unambiguous grammars that generated all of the strings produced by the original grammar also received full credit, regardless of how they fixed the problem.