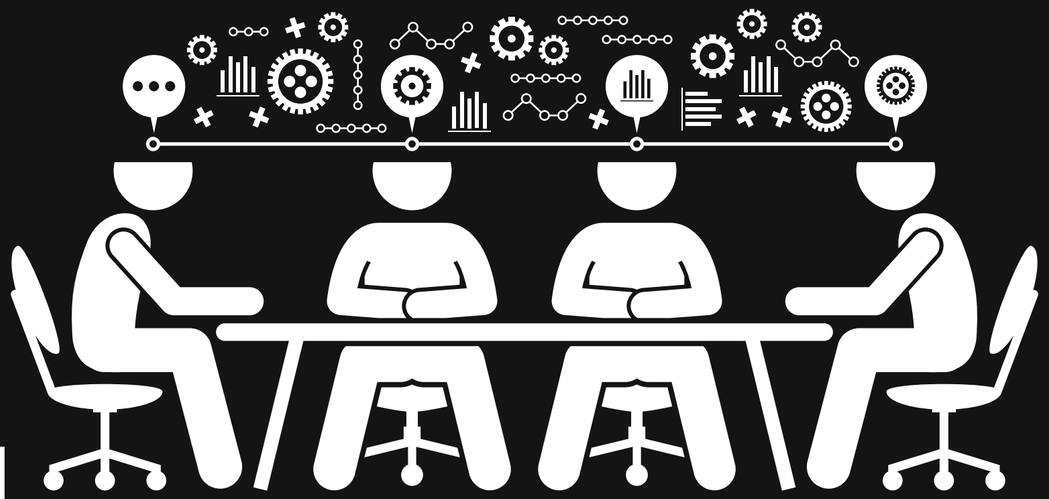
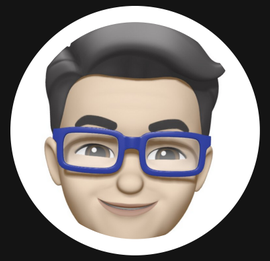
COPY CAT

ARCHIT MANEK, SHEBNA MATHEW, DEVINA RAITHATHA &
NIKHITA SINGH

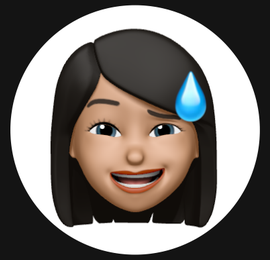
CS: 5500 - FOUNDATIONS OF SOFTWARE ENGINEERING

THE TEAM





ARCHIT MANEK



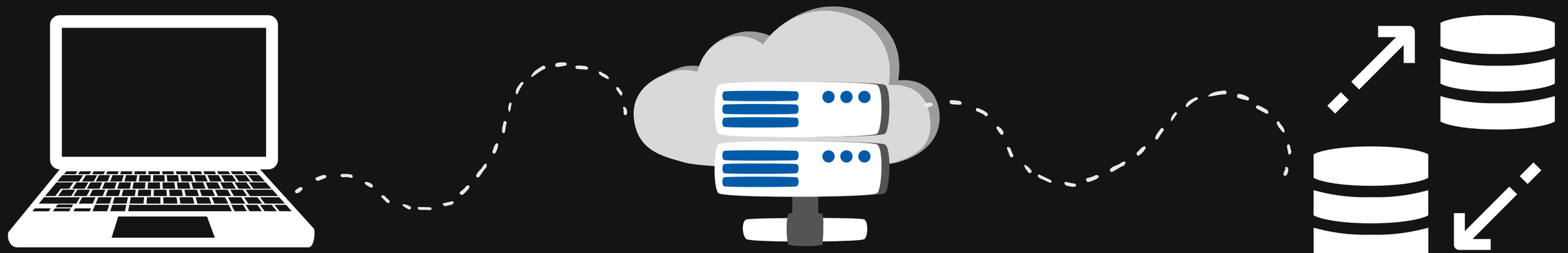
SHEBNA MATHEW

DEVINA RAITHATHA

NIKHITA SINGH

AGENDA

- HIGH-LEVEL ARCHITECURE
- PLAGIARISM DETECTION ALGORITHM
- INTERACTIVE PRESENTATION OF TOOL
- DESIGN EVOLUTION
- SOFTWARE DEVELOPMENT PRINCIPLES



CLIENT

SERVER

DATABASE

HIGH LEVEL ARCHITECURE

PLAGIARISM DETECTION ALGORITHM

PARSING AND TRAVERSING THE AST: BABEL

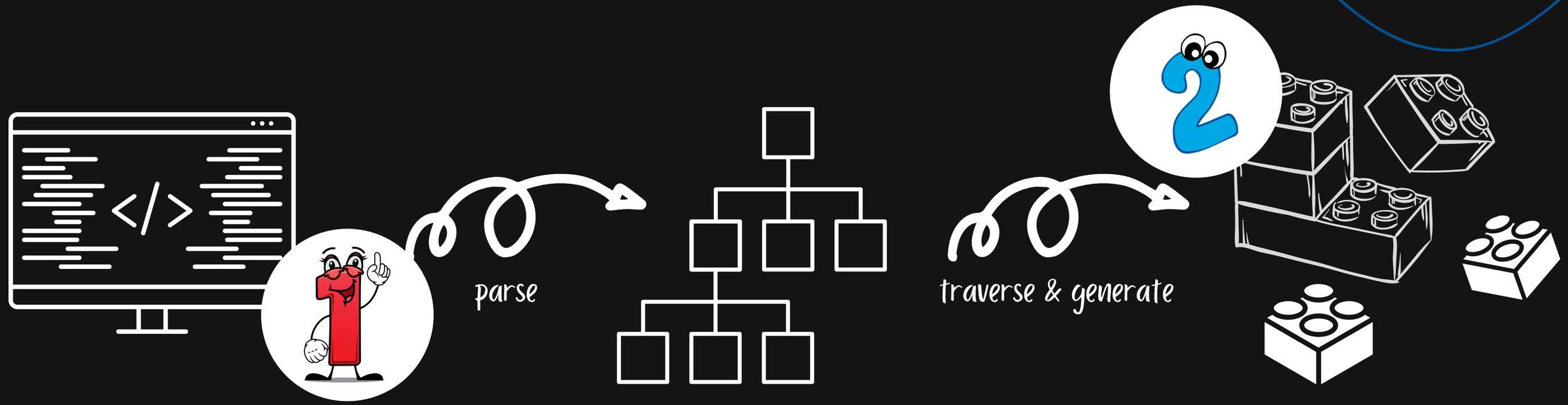
FOUR STRING REPRESENTATIONS OF THE SOURCE CODE

STRING - SIMILARITY COMPARISON

TOTAL PLAGIARISM %

PRE-PROCESSING & 4 STRINGS

PARSING: BUILDING AN AST



STUDENT FILE: SOURCE CODE

ABSTRACT SYNTAX TREE

STRING REPRESENTATIONS

FILE STRUCTURE:

“ProgramVariableDeclarationExpressionStatement”

IDENTIFIERS:

1f40fc92da241694750979ee6cf582f2d5d7d28e18335de05abc54d0560e0f5302860c652b
f08d560252aa5e74210546f369fbbbce8c12cfc7957b2652fe9a75eed55db3ffa1983455e1c
1b291f6d4d48009bbf43338657ac7a49631126140f126f0a95456b60535e3a7902acd712a
62044b445972a8b6c1e79c7cbbb80bc01f873bcc37e512b7da86476367769c932009fc1be
e59c929879f10ac89df541124db5010ae7297ec71db8f71a09adccd27c002f00fcfc93ca7e1
57105e7505f24d”

LITERALS:

“10HelloWorld”

TOKENS:

ProgramVariableDeclarationVariableDeclaratorIdentifierNumericLiter
alExpressionStatementCallExpressionMemberExpressionIdentifierIden
tifierStringLiteral

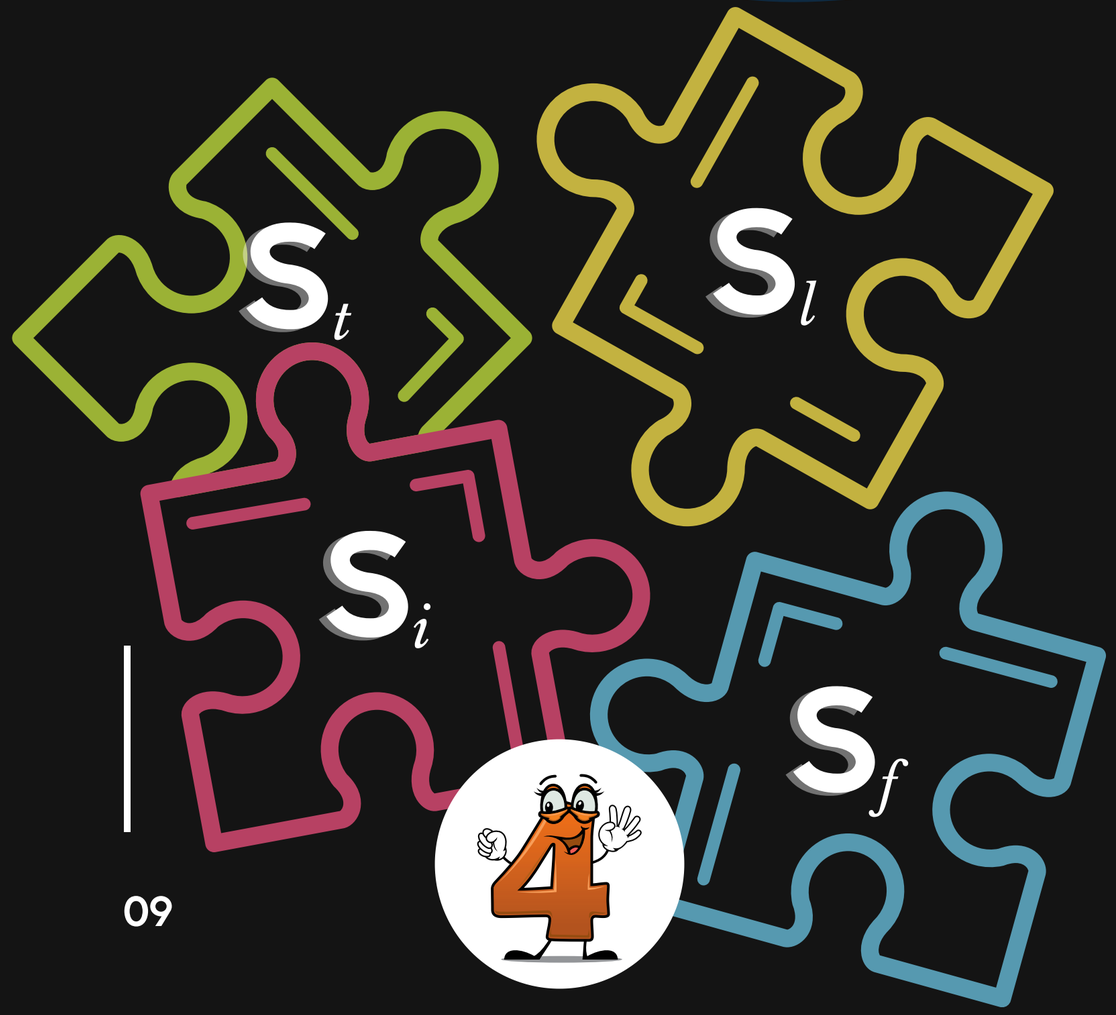
STRING SIMILARITY

ONCE WE HAVE THE FOUR STRINGS FOR EACH SOURCE CODE, WE GET THE DEGREE OF SIMILARITY BETWEEN TWO STRINGS.

FOR THIS WE USE A STRING COMPARE BASED ON THE DICE SIMILARITY COEFFICIENT WHICH IS A SIMPLE AND EFFECTIVE WAY TO CALCULATE A MEASURE OF THE SIMILARITY OF TWO STRINGS. THE OUTCOME IS A VALUE BOUNDED BETWEEN ZERO AND ONE THAT WE USE TO CALCULATE SIMILARITY PERCENTAGE.

STRING SIMILARITY

HOW SIMILAR ARE THESE TWO STRINGS?



$$\frac{S_f + S_i + 2 + S_l + 10 \cdot S_t}{14}$$

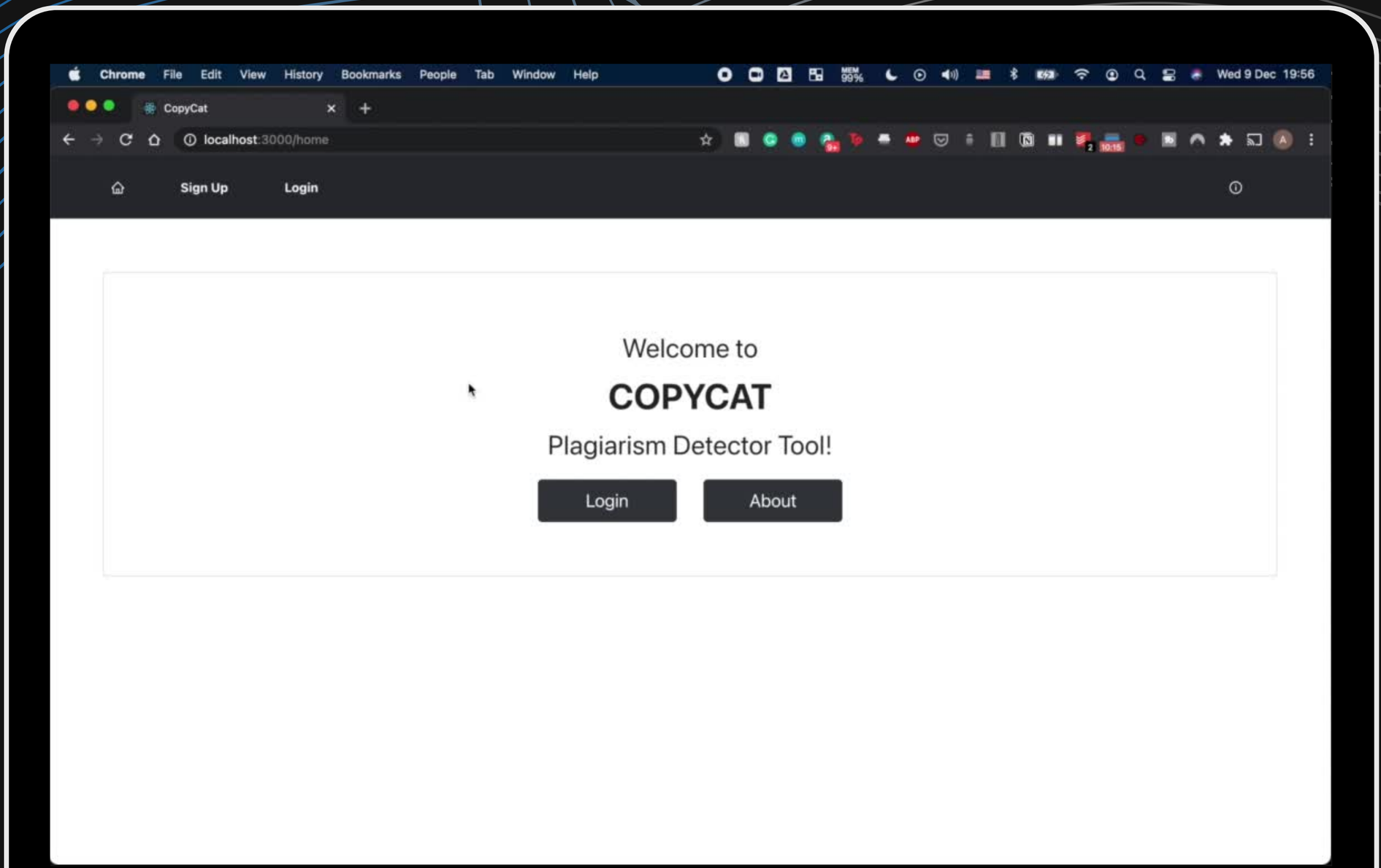
REPRESENTS THE TOTAL SIMILARITY AS A WEIGHTED AVERAGE OF THE FOUR SIMILARITIES: SF (FILE STRUCTURE), SI (IDENTIFIER), SL (LITERALS) AND ST (TOKENS).

STUDENT CODE FILE UPLOAD

DETECT PLAGIARISM

PLAGIARISM PERCENTAGE

HOW TO USE OUR TOOL


Chrome File Edit View History Bookmarks People Tab Window Help
Wed 9 Dec 19:56
CopyCat
localhost:3000/home
Sign Up
Login
Welcome to
COPYCAT
Plagiarism Detector Tool!
Login
About

DESIGN EVOLUTION

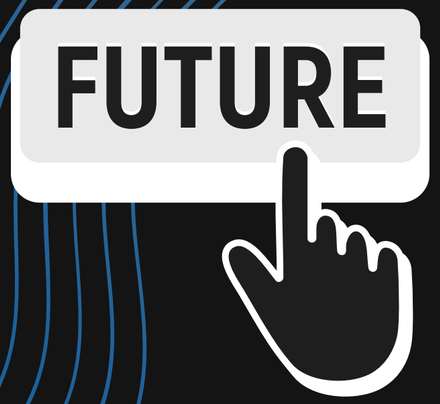
- WHAT PROGRAMMING LANGUAGES ARE SUPPORTED?
- WHAT LIBRARIES AND NPM PACKAGES CAN WE USE?
- HOW DO WE TRANSPOSE SOURCE CODE TO AN AST?

SOFTWARE DESIGN PRINCIPLES

WATERFALL METHODOLOGY

FACTORY PATTERN

VERSION CONTROL



SCOPE

MULTIPLE STUDENT COMPARISONS

MULTIPLE LANGUAGE SELECTION

SUPPORT COMPARISONS WITH ONLINE CODE

GENERATE & DOWNLOAD PLAGIARISM REPORT


THE END
Link to Slides: https://bit.ly/fse-slides