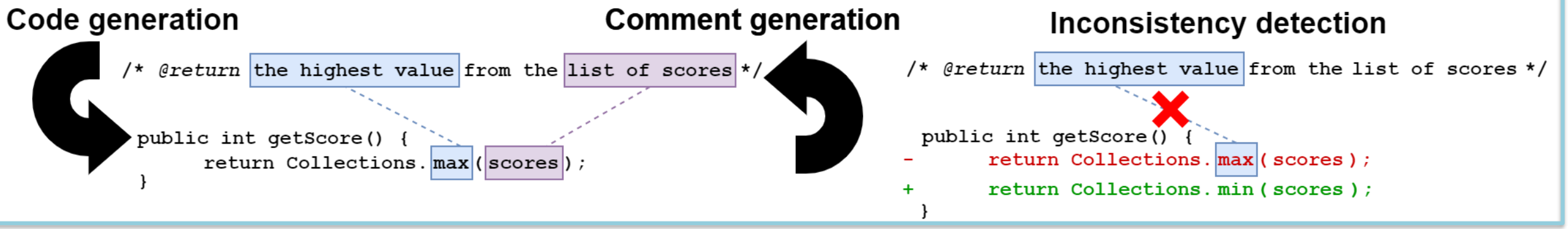
The University of Texas at Austin

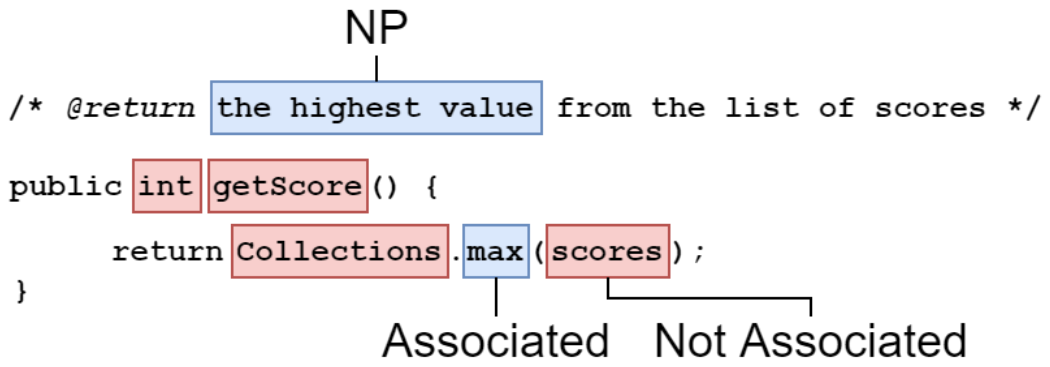
# Associating Natural Language Comment and Source Code Entities

**Sheena Panthaplackel (spantha@cs.utexas.edu), Milos Gligoric, Raymond J. Mooney, Junyi Jessy Li**
The University of Texas at Austin

**Motivation:** Learning to relate comment and code elements is critical to automated systems for generating code and comments and detecting inconsistent code and comments.




**Task:** Given a noun phrase (NP) in a comment, we classify the relationship between the NP and each candidate code token in the corresponding method as either associated or not associated.




Code and data available at: http://bit.ly/NLCodeAssociations

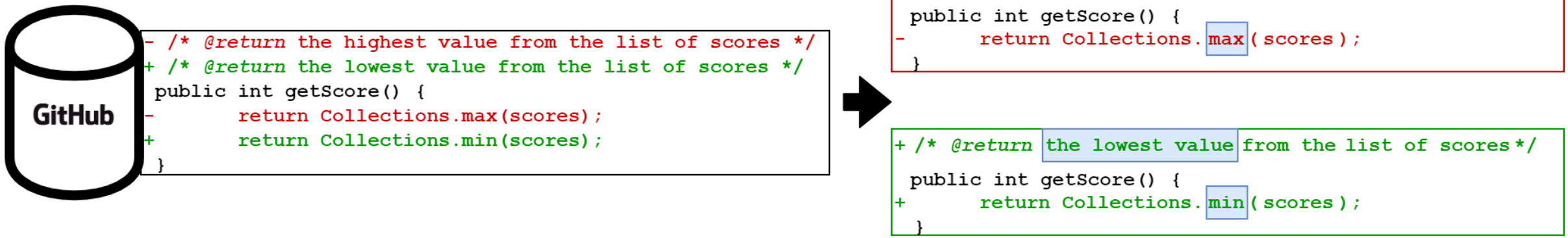
The University of Texas at Austin

# Associating Natural Language Comment and Source Code Entities

**Sheena Panthaplackel (spantha@cs.utexas.edu), Milos Gligoric, Raymond J. Mooney, Junyi Jessy Li**
The University of Texas at Austin

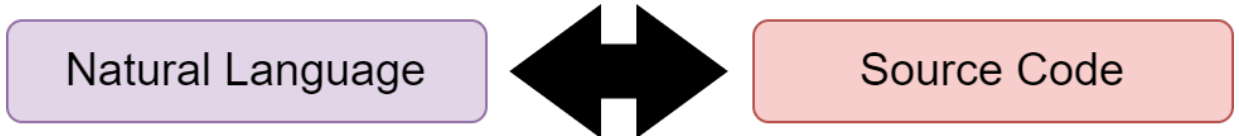
## Challenge #1: Obtaining labeled data

Noisy supervision: Mining simultaneous code/comment updates from GitHub commits




## Challenge #2: Reasoning across two distinct language representations

Salient features: Capturing characteristics of code, comments, and the relationship between them




Code and data available at: http://bit.ly/NLCodeAssociations