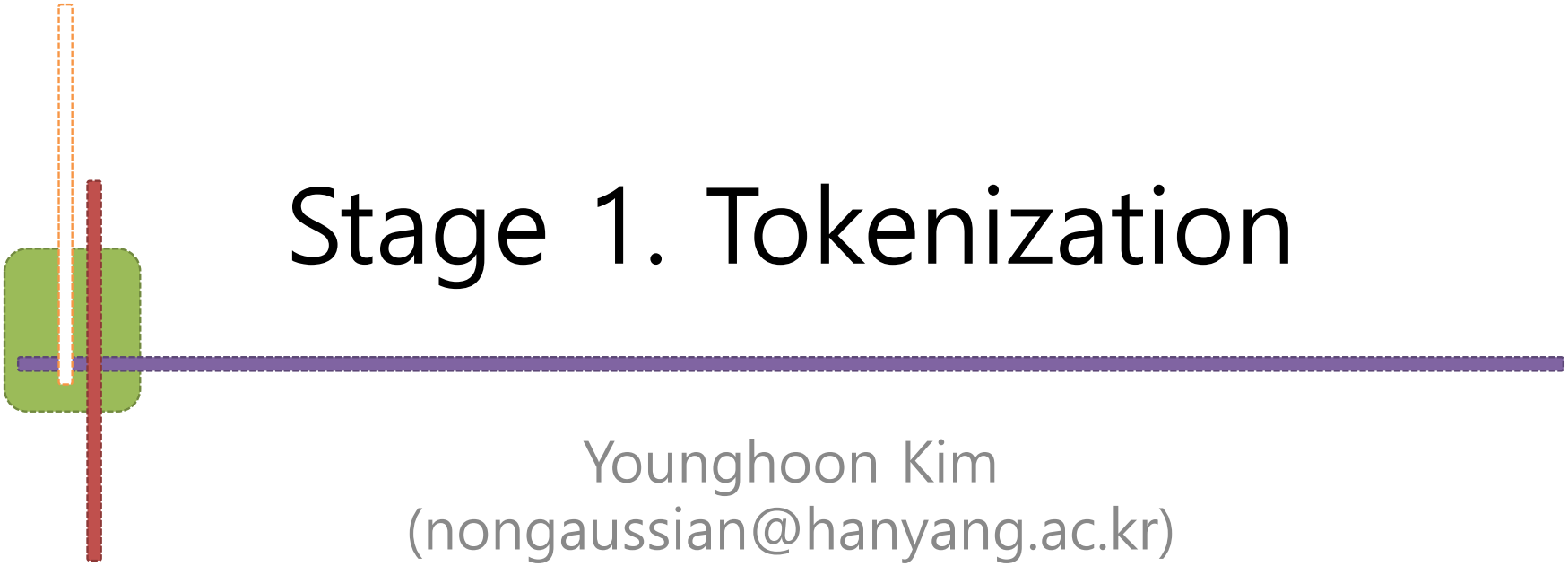

Stage 1. Tokenization

Younghoon Kim
(nongaussian@hanyang.ac.kr)

Sample Code

```
public class TinySETokenizer implements Tokenizer {
        private Analyzer analyzer = null;
        private PorterStemmer s = null;
        public void setup() {
                analyzer = new SimpleAnalyzer();
                s = new PorterStemmer();
        }
        public List<String> split(String text) {
                List<String> result = new ArrayList<String>();
                try {
                        TokenStream stream = analyzer.tokenStream(null, new StringReader(text));
                        while (stream.incrementToken()) {
                                result.add(stemString(
                                        stream.getAttribute(
                                        CharTermAttribute.class).toString()));
                        }
                        stream.close();
                } catch (IOException e) {
                        throw new RuntimeException(e);
                }
                return result;
        }
        public void clean() { analyzer.close(); }
        private String stemString(String word) { s.setCurrent(word); s.stem(); return s.getCurrent(); }
}
```