

```

package com.dlj.ir.analysis.be;

import java.util.Map;
import java.util.WeakHashMap;

/**
 * Bengali light stemmer- removes number, gender and case suffixes from nouns and
 * adjectives
 * Done by Ljiljana Dolamic (University of Neuchatel, www.unine.ch/info/clef/)
 *
 * @author Ljiljana Dolamic. Email: ljiljana.dolamic@unine.ch
 */
public class BengaliStemmerLight{

    /**
     * A cache of words and their stems
     */
    static private Map<String, String> cache = new WeakHashMap<String, String>();

    /**
     * A buffer of the current word being stemmed
     */
    private StringBuilder sb = new StringBuilder();

    /**
     * Default constructor
     */
    public BengaliStemmerLight() {
    }

    public String stem(String word) {
        String result = cache.get(word);

        if (result != null)
            return result;

        //
        sb.delete(0, sb.length());

        //
        sb.append(word);

        /* remove the case endings from nouns and adjectives */
        remove_case(sb);

        remove_article(sb);

        normalize(sb);

        result = sb.toString();
    }

```

```

        cache.put(word, result);
        return result;
    }

    private void remove_article(StringBuilder word) {
        int len = word.length() - 1;
        if (len > 5) {
            if (word.substring(len - 3, len + 1).equals("খানা")
                || word.substring(len - 3, len + 1).equals("খানি")
                || word.substring(len - 3, len + 1).equals("গুলো")) {
                word.delete(len - 3, len + 1);
                return;
            }
        } /* end if len >5 */
        if (len > 4) {
            if (word.substring(len - 2, len + 1).equals("য়োন")) {
                word.delete(len - 2, len + 1);
                return;
            }
        } /* end if len > 4 */
        if (len > 3) {
            if (word.substring(len - 1, len + 1).equals("টা")
                || word.substring(len - 1, len + 1).equals("টি")) {
                word.delete(len - 1, len + 1);
                return;
            }
        }
        return;
    }

    private void remove_case(StringBuilder word) {
        int len = word.length() - 1;

        if (len > 8) {
            if (word.substring(len - 6, len + 1).equals("য়দেরকে")) {
                word.delete(len - 6, len + 1);
                return;
            }
        } /* end if len >8 */
        if (len > 7) {
            if (word.substring(len - 5, len + 1).equals("দেরকে")) {
                word.delete(len - 5, len + 1);
                return;
            }
        } /* end if len >7 */
        if (len > 6) {
            if (word.substring(len - 4, len + 1).equals("দেরকে")
                || word.substring(len - 4, len + 1).equals("য়দের")) {
                word.delete(len - 4, len + 1);
                return;
            }
        }
    }

```

```

    } /* end if len >6 */
    if (len > 5) {
        if (word.substring(len - 3, len + 1).equals("দের")
            || word.substring(len - 3, len + 1).equals("য়রো")) {
            word.delete(len - 3, len + 1);
            return;
        }
    } /* end if len >5 */
    if (len > 4) {
        if (word.substring(len - 2, len + 1).equals("দর")||
            word.substring(len - 2, len + 1).equals("রো")
            ||word.substring(len - 2, len + 1).equals("য়র")) {
            word.delete(len - 2, len + 1);
            return;
        }
    }

    } /* end if len > 4 */
    if (len > 3) {
        if (word.substring(len - 1, len + 1).equals("র")) {
            word.delete(len - 1, len + 1);
            return;
        }
        if (word.substring(len - 1, len + 1).equals("ত")) {
            word.delete(len - 1, len + 1);
            return;
        }
        if (word.substring(len - 1, len + 1).equals("রা")) {
            word.delete(len, len + 1);
            return;
        }
        if (word.substring(len - 1, len + 1).equals("ক")) {
            word.setCharAt(len - 1, 'ড');
            word.delete(len, len + 1);
            return;
        }
    }
    } /* end if len > 3 */
    if (len > 2) {
        if (word.charAt(len) == 'ঢ'
            || word.charAt(len) == 'ঝ'
            || word.charAt(len) == 'ঞ') {
            word.delete(len, len + 1);
            return;
        }
    }
    } /* end if len > 2 */
    return;
}

private void normalize(StringBuilder word) {
    int len = word.length() - 1;
    if (len > 4) {
        if (word.substring(len - 2, len + 1).equals("নি") {
            word.delete(len - 2, len + 1);

```

```
        return;
    }
} /* end if len > 4 */
if (len > 3) {
    if (word.substring(len - 1, len + 1).equals("नी")) {
        word.delete(len - 1, len + 1);
        len-=2;
    }
} /* end if len > 3 */
if (len > 2) {
    if (word.substring(len, len + 1).equals("ृ")
        || word.substring(len, len + 1).equals("ृ")) {
        word.delete(len, len + 1);
        return;
    }
} /* end if len > 2 */
return;
}
}
```