sani@ sharif.edu   hesami@ce.sharif.edu

:   .

.

.

.

.

.

.

.   :

.

.

(...   )

.

.

.   [ ].

.   .   .

.   :

:   ●

.[   ]

.   .

:   ●

.

"   "

'teacher'   " " " "

.   'er' 'teach'

s/es

ly

s   'o'

es

goe + s   goes

go + es

[   ]

word	Main stem	Find stem
going	go	goe
being	be	bee

[ ]

.[ ]

.[ ]

bee   being

---

[1] word net

[2] Morphology

[3] Krovetz

ing   bee

be

(, , , }+	
+	
{ , , , , , }+ +	
{ , , , , }+ + +	
{ , , , , }+ + +	
{ , , , , , }+ +	
{ , , , , , }+ +	
{ , , , , , }+ +	
{ , ,""}+ +	
{ , ,""}+ +	
, , , }+ { , ,	

(morphological)

[ ]

" "   :) (:)   (

(   )

Find (w)	.		w
Prefix (W, prf)	W	pfr	
	.		
Postfix (W, pof)	W	pof	
	.		
RePrefix (W, prf)	.	w	prf
RePostfix (W, pof)	.	w	prf

```
Word: The word for stemming.
Result: result of stemming.
//if the word is in dictionary don't need stemming
W1 = Word;
If find(Word)
Result = Word;
//This part if for werb begin with "می"
//مضارع اخباري و ماضي التزامي
If (PreFix (Word,"می")){
          Reprefix(Word,"می");
          Result +="می";
          If  find(Word){
                    Result += Word;
                    Return;        }
If(PostFix(Word,{ "م", "ي", "د", "يم","يد", "ند" })){
          RePostFix(Word, { "م", "ي", "د", "يم","يد", "ند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "م", "ي", "د", "يم","يد", "ند" };
Return;}}
Word = W1;
//امر و ماضي ساده
If (PreFix (Word,"ب")){
          Reprefix(Word,"ب");
          Result +="ب";
          If  find(Word){
                    Result += Word;
                    Return;        }
If(PostFix(Word,{ "م", "ي", "د", "يم","يد", "ند" })){
          RePostFix(Word, { "م", "ي", "د", "يم","يد", "ند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "م", "ي", "د", "يم","يد", "ند" };
Return;}}
//امر و ماضي ساده منفي
If (PreFix (Word,"ن")){
          Reprefix(Word,"ن");
          Result +="ن";
          If  find(Word){
                    Result += Word;
                    Return;}
→
```

```
→
If(PostFix(Word,{ "م", "ي", "د", "يم","يد", "ند" })){
          RePostFix(Word, { "م", "ي", "د", "يم","يد", "ند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "م", "ي", "د", "يم","يد", "ند" };
Return;}}
Word = W1;
//ماضي ساده
If(PostFix(Word,{ "م", "ي", "يم","يد", "ند" })){
RePostFix(Word, { "م", "ي", "يم","يد", "ند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "م", "ي", "يم","يد", "ند" };
Return;}
Word = W1;
//ماضي نقلي
If(PostFix(Word,{ "ه ايد", "ه ايم", "ه است", "ه ا ي", "ه ام", "ه اند" })){
RePostFix(Word, { "ه ايد", "ه ايم", "ه است", "ه ا ي", "ه ام", "ه اند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "ه ايد", "ه ايم", "ه است", "ه ا ي", "ه ام", "ه اند" };
Return;}
Word = W1;
//ماضي بعيد
If(PostFix(Word,{ "ه بوديد", "ه بوديم", "ه بود", "ه بودي", "ه بودم", "ه بودند" })){
RePostFix(Word, { "ه بوديد", "ه بوديم", "ه بود", "ه بودي", "ه بودم", "ه بودند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "ه بوديد", "ه بوديم", "ه بود", "ه بودي", "ه بودم", "ه بودند" };
Return;}
Word = W1;
//ماضي التزامي
If(PostFix(Word,{ "ه باشيد", "ه باشيم", "ه باشد", "ه باشي", "ه باشم", "ه باشند" })){
RePostFix(Word, { "ه باشيد", "ه باشيم", "ه باشد", "ه باشي", "ه باشم", "ه باشند" });
          If(find(Word){
                    Result += Word + "+";
Result += { "ه باشيد", "ه باشيم", "ه باشد", "ه باشي", "ه باشم", "ه باشند" };
Return;}
 Word = W1;
If (prefix(Word,"ن")
{//Remove the "ن" from prefix and check all part for it  according the table 1}
```

+ +	+	
+ +	+	

[1] R. Krovetz, 1993: **"Viewing morphology as an inference process"**, in R. Korfhage et al., Proc. 16th ACM SIGIR Conference, Pittsburgh, June 27-July 1, 1993; pp. 191-202.

[2] J. P. Callan, M. Connell, A. Du. "**Automatic discovery of language models for text databases**", *ACM SIGMOD International Conference on Management of Data*, pages 479-490, 1999.

[3] K. Aas and L. Eikvil, "*Text Categorisation: A Survey*", http://citeseer.nj.nec.com/ , aas99text.html, 1999.

[4] M. F. Porter. "**An algorithm for suffix stripping**", *Program*, 14(3):130t137, 1980.

[5] E.M Voorhees, "**using wordnet for text retrieval**", WordNet an electronic lexical Database, Page 285-303, MIT Press.

[6] Kazem Taghva, Russell Beckley, Mohammad Sadeh, "**A Stemming Algorithm for the Farsi Language"**. ITCC (1): 158-162, 2005.

کراوتز	۲۲۳	۲۱۰	۱۳	۹۴,۱۷
کراوتز۲	۲۲۳	۲۱۹	۴	۹۸,۲۰