

# WEB-DEX

## 1. BASIC INFORMATION

*1.1 Lexicon type (wordform, explanatory, terminological lexicon, wordnet, etc),*  
WEB-DEX is an explanatory dictionary based on the 1996 edition of the standard explanatory dictionary of Romanian published by the Romanian Academy

*1.2 Representation of the lexicon (flat files, database, markup)*  
WEB-DEX is distributed as an XML file.

*1.3 Character encoding*  
The characters have been encoded in UTF8

## 2. ADMINISTRATIVE INFORMATION

*2.1 Contact person*  
Name: Dan Tufis,  
Address: Calea 13 Septembrie, no. 13, 050711  
Affiliation: Research Institute for Artificial Intelligence, Romanian Academy  
Position: Director  
Telephone: +4021 3188103  
Fax: +40 21 3188142  
e-mail: [tufis@racai.ro](mailto:tufis@racai.ro)

*2.2 Delivery medium (if relevant; description of the content of each piece of medium)*  
The resource will be uploaded on the MetaShare platform as an archive with the structure as described under 3.1.

*2.3 Copyright statement and information on IPR*  
The resource is free license-based for research purposes and fee license-based for commercial purposes

## 3. TECHNICAL INFORMATION

*3.1 Directories and files*  
The archive contains an XML file (dex.utf8.xml).

*3.2 Data structure of an entry*  
The structure of an entry in WEB-DEX is exemplified below:

```
<entry id="JACHETĂ">
  <hw>JACHETĂ</hw><stress>JACH`ETĂ</stress>
  <alt><brack> <gram>nominativ_feminin_singular_indefinit</gram>
    <orth>jachetă;</orth></brack>
  <brack> <gram>nominativ_feminin_plural_indefinit</gram>
    <orth>jachete</orth></brack></alt>
  <pos>substantiv</pos>
  <gen>feminin</gen>
  <struc>
    <def>Haină ( tricotate) femeiască încheiată în față, care acoperă partea de sus a corpului și
      care se poartă peste bluză sau peste rochie </def>
    <struc type="Sec">
      <def>Haină bărbătească de ceremonii, croită pe talie, lungă până aproape de genunchi.
        </def>
    </struc>
  </struc>
</entry>
```

</struc>  
<etym>Din limba <lang>fr.</lang>jaquette</etym>  
</entry>

### 3.3 Lexicon size (nmb. of lexical items, KB occupied on disk)

54.222 entries, requiring 44 MB of hard disk

## 4. CONTENT INFORMATION

### 4.1 The natural language(s) of the lexicon

The language of the lexicon is Romanian.

### 4.2 Entry Type

Each entry in the printed dictionary is represented as described at item 3.2.

### 4.3 Attributes and their values

See section 3.2:

### 4.4 Coverage of the lexicon

The lexical stock covers the basic general language vocabulary of Romanian

### 4.5 Intended application of the lexicon

General lexicographic studies, NLP applications, IR tasks, etc.

### 4.6 POS assignment

The part of speech assignment is the one in the printed Explanatory Dictionary of Romanian

### 4.7 Reliability (automatically/manually constructed)

Manually constructed based on the printed form.

## 5. RELEVANT REFERENCES AND OTHER INFORMATION

1. Tufiş, D., Rotariu, G., Barbu A.M. "Data Sampling, Lemma Selection and a Core Explanatory Dictionary of Romanian". In Proceedings of the 5th International Workshop on Computational Lexicography COMPLEX, Pecs, Ungaria, 1999, pp. 219-228
2. Erjavec, T., Tufiş, D., Varadi T. "Developing TEI-Conformant Lexical Databases for CEE Languages". In Proceedings of the 4th International Workshop on Computational Lexicography COMPLEX, Pecs, Ungaria, 1999, pp.205-210.
3. Dimitrova L., Erjavec T., Ide N., Kaalep H.J., Petkevic V., Dan Tufiş: Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages, COLING, Montreal 1998