“ALKHALILDWS”: AN ARABIC DICTIONARY WRITING SYSTEM RICH IN LEXICAL RESOURCES

MOHAMMED REQQASS[1], ABDELHAK LAKHOUAJA [1] AND MOHAMAD BEBAH [2]

16 – 17 October 2019
ICALP 2019, Nancy France

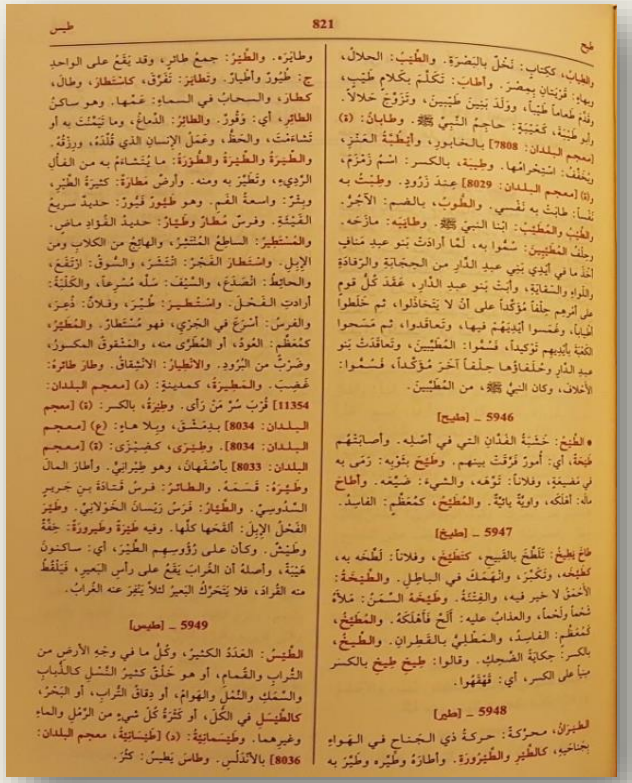
Mohammed Reqqass[1], Abdelhak Lakhouaja [1] and Mohamad Bebah [2]
[1] Faculty of Sciences, Mohamed First University, Av Med VI BP 717, Oujda 60000, Morocco
reqqass.mohammed@gmail.com, abdel.lakh@gmail.com
[2] Arab Center for Research and Policy Studies, Doha, Qatar
mohamad.bebah@dohainstitute.org

OUTLINE

- Introduction
- Dictionary Writing System
- "AlkhalilDWS": An Arabic DWS rich in lexical resources
- Conclusion

Introduction

- A dictionary is a support that list the words of one or more specific languages, often listed alphabetically
- The process of making dictionary:
  - Identification of lexical entries
  - Editing lexical entries
  - Publication of lexical entries
- “AlkhalilDWS” assist the editor to make their Arabic dictionary project



Dictionary Writing System

Software for writing and producing a dictionary. It meet the following needs:

- write and edit dictionary entries
- control of the quality entries and ensuring their consistency
- exploitation of the available linguistic resources
- production of the dictionary in several forms: paper, electronic…
- exchange with other applications
- make collaborative group work

Dictionary Writing System

Dictionary Writing System	Description
TshwaneLex	a commercial writing system of monolingual dictionaries, bilingual dictionaries and multilingual dictionaries
Matapuna	an open source dictionary writing system, used to build a dictionary of Maori language
Glossword	is a system for writing and publishing multi-language dictionaries. It is a web application, open source.
ABBYY Lingvo Content	a dictionary writing system intended for compiling dictionaries, glossaries, encyclopedias, and other types of reference materials.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Nature of DWS
  - Based on a server– client architecture ;
- Data Base of DWS
  - Allows multi-user to work on the same dictionary-making project from any location ;
  - The data must be stored in a structured and extensible database;

"AlkhalilDWS": An Arabic Dictionary Writing System

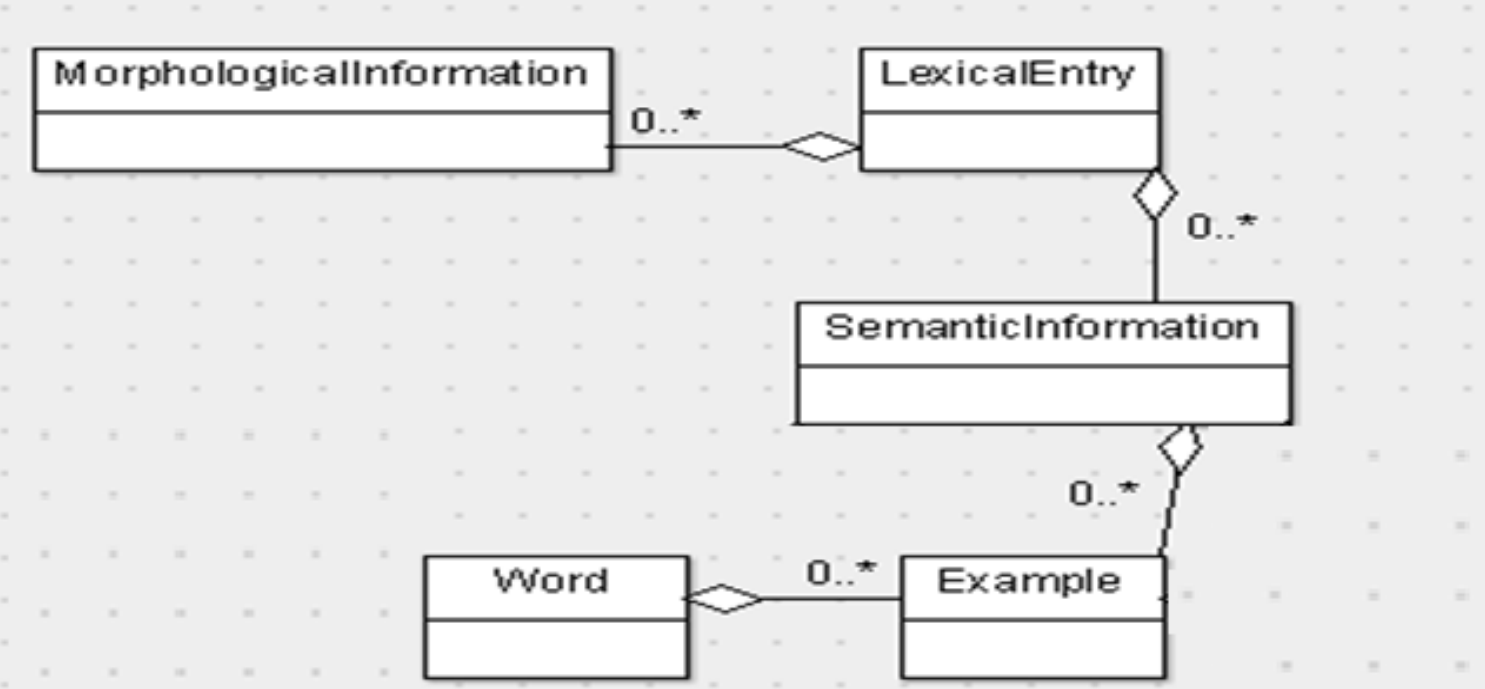
- Morphological analyzer
  - is analysis of structure of the word forms, it provides information about: word category (nouns, verb, adj, etc.), root, lemma, gender, etc.
  - find new lexical entries,
  - add grammatical information to lexical entries,
  - improve the search for a lexical entry in the dictionary.

"AlkhalilDWS": An Arabic Dictionary Writing System

- The spell checker
  - a software feature that checks for misspellings in a text;
  - errors categories : reading errors, hearing errors ,touch-typing errors, morphological errors, editing errors

"AlkhalilDWS": An Arabic Dictionary Writing System

- Lexical Markup Framework
  - model that provides a common standardized framework for the construction of natural language processing lexicons



"AlkhalilDWS": An Arabic Dictionary Writing System

- Dictionary Structure
- Lexical resources in "AlkhalilDWS"
- Operating "The Dictionary of the Modern Arabic Language"
- Profiles in "AlkhalilDWS"
- Identification of lexical entries
- Edition of lexical entries
- Publishing of lexical entries

"AlkhalilDWS": An Arabic Dictionary Writing System

- Dictionary Structure
  - **macrostructure**: editor defines the global order of the lexical entries in the dictionary;
  - **microstructure**: editor defines the internal structure of the lexical entry and the information it contains;

"AlkhalilDWS": An Arabic Dictionary Writing System

- Lexical resources in "AlkhalilDWS"
  - "**The interactive dictionary of the Arabic language**": an interactive open source web application based on "Alwassyt" dictionary.
  - "**Almustakshif dictionary**": the first Arabic lexical encyclopedia of the dictionary "AlMoheet"
  - **"The Dictionary of the Modern Arabic Language"**: dictionary targeting to cover classical Arabic words still in usage, as well the new words used in different Arab countries.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Lexical resources in "AlkhalilDWS"
  - "The interactive dictionary of the Arabic language": an interactive open source web application based on "Alwassyt" dictionary.
  - "Almustakshif dictionary": the first Arabic lexical encyclopedia of the dictionary "AlMoheet"
  - **"The Dictionary of the Modern Arabic Language"**: dictionary targeting to cover classical Arabic words still in usage, as well the new words used in different Arab countries.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Operating "The Dictionary of the Modern Arabic Language":
  - convert the document format of the dictionary to a text file,
  - keep only the lexical entries (remove introduction, index);
  - identify and build a mapping root, headword of lexical entries and the content of lexical entries;
    - The root is extracted from the line that start with a number followed by the character "-" and isolated Arabic characters. We keep the isolated Arabic characters only.
    - The headword is extracted from the line that does not start with a number or that start with the character "•". We keep the word that appears before the character ":" and the character "]".

“AlkhalilDWS”: An Arabic Dictionary Writing System

- Operating “The Dictionary of the Modern Arabic Language”:
  - The content of lexical entries is extracted from the line that does not contain root and from the line that contains a headword. If the line contains a headword, we keep only the words appearing before the first occurrence of the character ":";
    - extract meaning, semantic field and example from the content of lexical entry;
    - the semantic is the word that appears between the first occurrence of the character ")" and the first occurrence of the character "(";
    - the examples appear between the character “"", each example is separated by the character "-” or the character “° ”;
    - the meaning is the set of words that do not represent the semantics or the examples.
- normalize the writing of the root by removing the space with the isolate Arabic character and replace the characters “أ”, “إ”, “ا” by the character “ء”;
- normalize the writing of the lemma by analyzing them with the morphological analyzer Alkhalil2 and keeping only the result that corresponds to the lexical entries.

“AlkhalilDWS”: An Arabic Dictionary Writing System

- Operating “The Dictionary of the Modern Arabic Language”:

149 - أ س د

أَسَد [مفرد]: ج آساد وأُسْد وأُسُد وأُسُود، مؤ أَسَدة:

1 - (حن) حيوان مفترس شديد الضراوة من فصيلة السِّنّوريّات ورُتْبة اللَّواحم، يشمل الذكر والأنثى ويطلق على الأنثى أسدة ولَبُؤَة، وله في العربيّة أسماء كثيرة أشهرها الليث والضيغم والغضنفر والضرغام "رأيت أسدًا- هذا الشِّبل من ذاك الأسد: يشبه الابن أباه في صفاته- *أَسَد عليّ وفي الحروب نعامةٌ*" ° أسد الله: حمزة بن عبد المطلب رضي الله عنه- بين فكَّيّ الأسد: في خطر، في مأزق- حِصَّة الأسد: الجزء الأكبر.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Operating "The Dictionary of the Modern Arabic Language":

Element	Translation	Content
Root	A s d	أ س د
normalized root	a s d	ءسد
Lemma	Lion (assad)	أَسَد
normalized lemma	lion	أَسَد
Meaning	Very predatory animal. wild member of the felidae family. It includes males and females. The female is called lioness. it has many names in Arabic, including "leith", "Ghadanfar", "Dirgham" etc..	حيوان مفترس شديد الضراوة من فصيلة السِّنّوريّات ورُتْبة اللَّواحم، يشمل الذكر والأنثى ويطلق على الأنثى أسدة ولَبُؤَة، وله في العربيّة أسماء كثيرة أشهرها الليث والضيغم والغضنفر والضرغام
Semantic	Animal	حيوان
Examples	• I saw a lion • This cub is from that lion • A lion against me and soft in the war (proverb) • Lion of Allah: Hamza ibn Abdul-Muttalib (person) • Between lion's jaws: in danger, in trouble • The lion's share: the major share of something	• رأيت أسدًا • هذا الشِّبل من ذاك الأسد: يشبه الابن أباه في صفاته • *أَسَد عليّ وفي الحروب نعامةٌ* • أسد الله: حمزة بن عبد المطلب رضي الله عنه • بين فكَّيّ الأسد: في خطر، في مأزق • حِصَّة الأسد: الجزء الأكبر

"AlkhalilDWS": An Arabic Dictionary Writing System

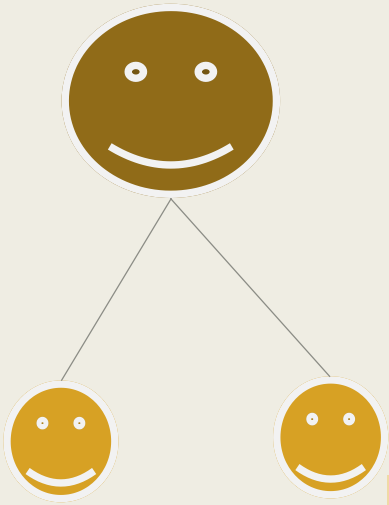
- Lexical resources in "AlkhalilDWS"

Dictionary	root	lemma	meaning	example	semantic field
the interactive dictionary of the Arabic language	7081	133591	133594	11643	247
the modern Arabic dictionary	5739	32258	91874	30909	35
almustakshif dictionary	5386	20112	26561	0	37

“AlkhalilDWS”: An Arabic Dictionary Writing System

- Profiles in “AlkhalilDWS”:

chief editor: It allows the user to create his own working group, to identify lexical entries, to assign lexical entries to the members of his group, to approve the works submitted by editors, to export the dictionary in XML.



administrator: manage the accounts of chief editors.

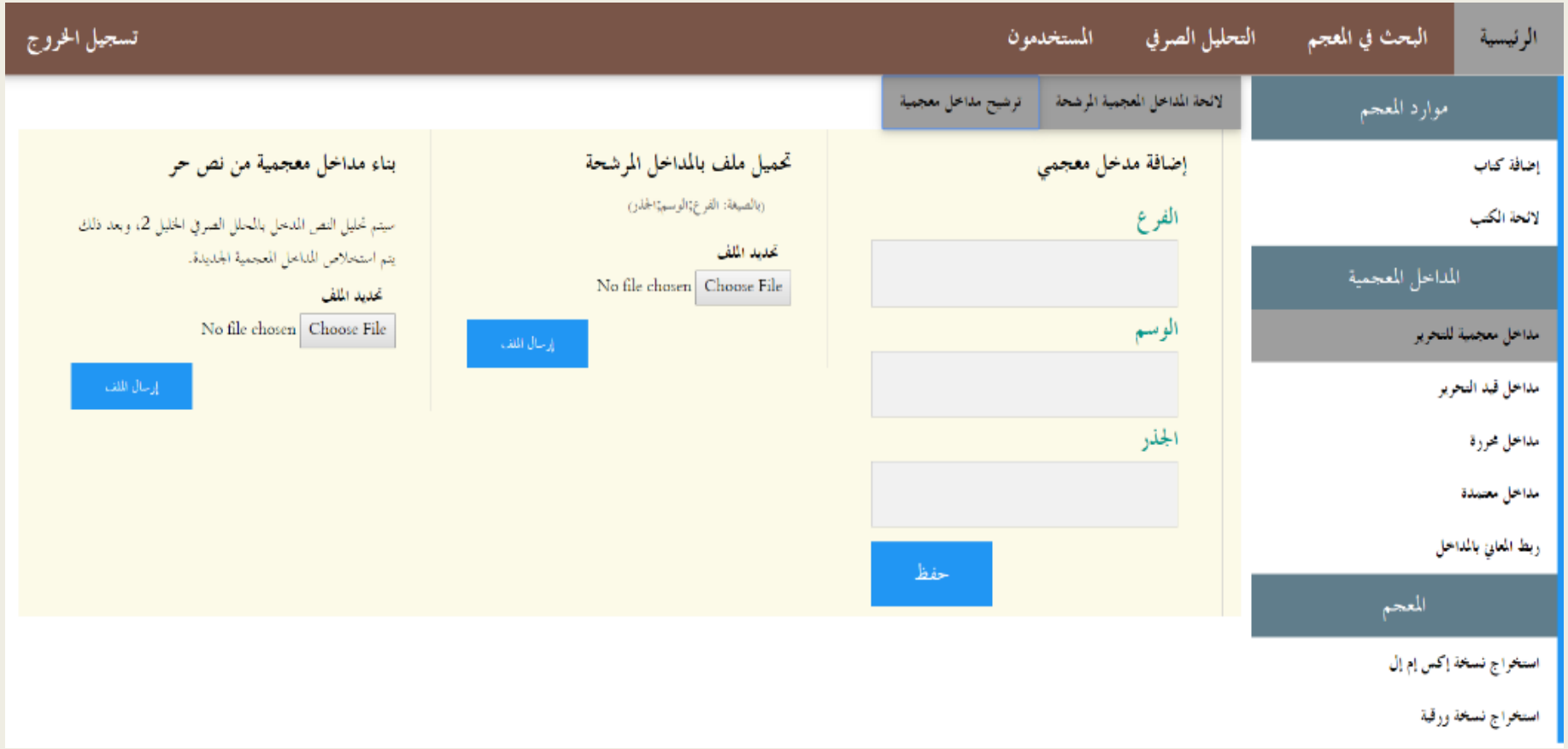
editor: It allows the user to edit entries affected to him.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Identification of lexical entries
  - **manual entering**: the editor enters the words one by one. This is done by filling in the following information: lemma, type, and root.
  - **importation of lemmas**: the editor imports a list of words from a text file. Each line in the file contains the morphological information of word: lemma; type; root.
  - **automatic extraction of lemmas**: the editor can download a text file. "AlkhalilDWS" analyzes the words of the text and extract the new lemmas. Every lemma is linked with its morphological information.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Identification of lexical entries

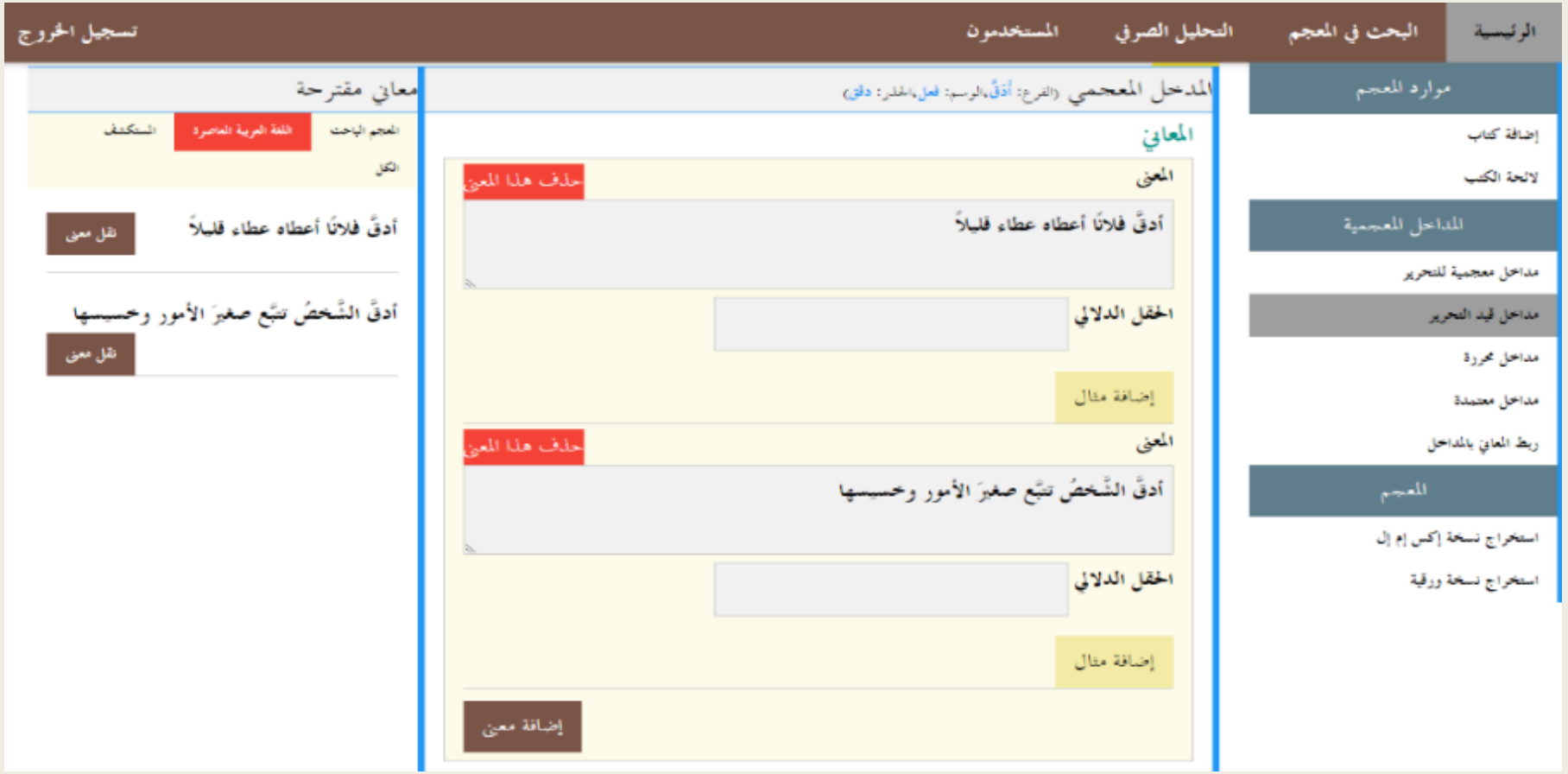


“AlkhalilDWS”: An Arabic Dictionary Writing System

- Identification of lexical entries

"AlkhalilDWS": An Arabic Dictionary Writing System

- Edition of lexical entries

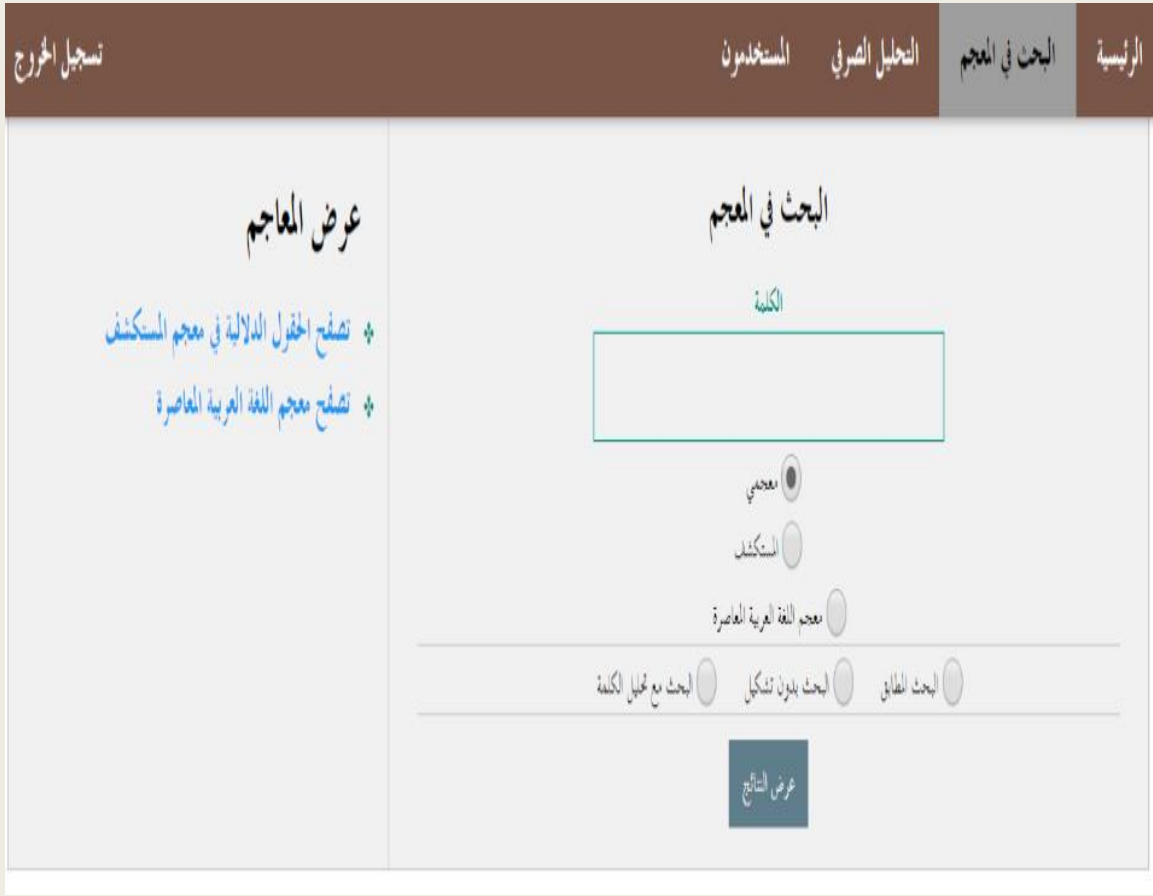


"AlkhalilDWS": An Arabic Dictionary Writing System

- Publishing of lexical entries

the interactive version: "AlkhalilDWS" offers an interactive version of the dictionary. It allows the chief editor to search in his produced dictionary with the following options:

- **search with exact matching**: search the exact word in the lexical entries;
- **search with exact matching and ignore the diacritics**: search the lexical entries that correspond to the exact word ignoring the diacritics;
- **search after analyzing word**: search the lexical entries that correspond to the lemma of the search entry.



“AlkhalilDWS”: An Arabic Dictionary Writing System

- Publishing of lexical entries

the interactive version: "AlkhalilDWS" offers an interactive version of the dictionary. It allows the chief editor to search in his produced dictionary with the following options:

- **search with exact matching**: search the exact word in the lexical entries;
- **search with exact matching and ignore the diacritics**: search the lexical entries that correspond to the exact word ignoring the diacritics;
- **search after analyzing word**: search the lexical entries that correspond to the lemma of the search entry.

"AlkhalilDWS": An Arabic Dictionary Writing System

- Publishing of lexical entries

XML format: The chief editor can export his produced dictionary in XML format. This format respects the standard LMF. Bellow an example of some entries edited with "AlkhalilDWS"

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<Dictionary>
<entry id="1887">
<feat att="partOfSpeach" val="اسم"/>
<root>
<feat att="writtenForm" val="ءخر"/>
</root>
<lemme>
<feat att="writtenForm" val="آخَر"/>
</lemme>
<meaning>
<definition>
<feat att="text" val="مختلف، مغاير أو بمعنى غيره"/>
</definition>
</meaning>
<meaning>
<definition>
<feat att="text" val="أحد شيئين يكونان من جنس واحد"/>
</definition>
</meaning>
</entry>
<entry id="2276">
<feat att="partOfSpeach" val="فعل"/>
<root>
<feat att="writtenForm" val="ءون"/>
</root>
<lemme>
<feat att="writtenForm" val="آنَ"/>
</lemme>
<meaning>
<definition>
<feat att="text" val="صَبَّه  :أنَ الماء."/>
</definition>
</meaning>
<meaning>
<definition>
<feat att="text" val="ماحبُها  :أنا خالَ هذا الفرس."/>
</definition>
</meaning>
</entry>
</Dictionary>
```

Conclusion

- "AlkhalilDWS" meets the requirements of Arabic lexicography. It offers many features and lexical resources that make the construction and updating of Arabic dictionaries easier.
- We have adopted the LMF standard to ensure the exchange and communication between produced dictionaries and any tool that complies with this standard

➔ In the future, we plan to enrich "AlkhalilDWS" with new features as well as extracting good examples from the corpus, allowing the chief editor to define the structure of the lexical entry.

Thank you for your attention