

Implementation of ARSEM database and interfaces

Swithun Crowe

Application Developer
Research Computing
University of St Andrews
cs2@st-andrews.ac.uk

26 September 2013



1 Research Computing at St Andrews

2 Technologies Used

- Text Encoding Initiative
- Apache Solr
- W3C XForms

3 TEI Documents and Interfaces

- Interfaces
- Dictionary Entries
- Poems



Research Computing at St Andrews

www.st-andrews.ac.uk/itsupport/academic/research

- IslamAnatolia
 - ▶ islam-anatolia.ac.uk
 - ▶ P.I. Dr Andrew Peacock
 - ▶ Studies the transformation of Anatolia from a Christian to a predominantly Muslim society over the period c. 1100 to 1500AD.
- Publishing the Philosophical Transactions
 - ▶ arts.st-andrews.ac.uk/philosophicaltransactions
 - ▶ P.I. Dr Aileen Fyfe
 - ▶ Investigating the challenges and opportunities of scholarly publishing over the past 350 years.
- University Library Digital Collections
 - ▶ arts.st-andrews.ac.uk/digitalhumanities
 - ▶ Uses Islandora (Drupal + Fedora Commons)



Text Encoding Initiative

www.tei-c.org

- XML vocabulary widely used in humanities for marking up prose, poetry, drama, linguistic texts, dictionaries, manuscripts etc.
- TEI document headers contain rich metadata - provenance, bibliographic and technical information
- Easily edited using schema-aware software, e.g. Oxygen
- Transformed to HTML using XSLT



Apache Solr

lucene.apache.org/solr

- Very fast search engine, providing full text search, faceted searches, highlighting of search terms etc.
- Indexes XML documents (using XSLT), as well as common formats, e.g. Word and PDF
- Handles different languages, e.g. for stemming, ASCII folding
- Used to find poems using lemmata/senses, lemmata/senses present in poems, phraseology (consonant roots occurring in proximity)



W3C XForms

www.w3.org/TR/xforms

- XML language for describing form interfaces
- Form data are XML documents, rather than a flat list of name/value pairs
- This provides a more efficient workflow - XML documents remain XML documents throughout the process
- Using XSLTForms (www.agencexml.com/xsltforms) - a cross browser, client side implementation of XForms



Interfaces

- XForms interfaces for dictionary entries, poems and authors
- Bibliography managed by Zotero and can be uploaded to ARSEM database
- Users will be able to view poems and dictionary entries, and use links between these



Dictionary Entries

TEI and XForms showing lemma, grammar information

Lemma

Transliterated

Arabic

Foreign root

Root

Transliterated

Arabic

Grammar group

Part of speech

Subcategory

Vowel

Syntactic note

```

<form type="lemma">
  <orth lang="ara-Latn-x-1c">bakā</orth>
  <orth lang="ara">بَكَى</orth>
  <orth lang="" type="foreign"/>
  <ref n=""/>
</form>
<gramGrp>
  <pos>100</pos>
  <subc>l</subc>
  <gram type="vowel">v_i</gram>
  <gramGrp type="mandar">
    <gramGrp>
      <gram lang="ara">بَكَى</gram>
      <gram lang="ara-Latn-x-1c">bukā</gram>
    </gramGrp>
    <gramGrp>
      <gram lang="ara">بَكَى</gram>
      <gram lang="ara-Latn-x-1c">bukan</gram>
    </gramGrp>
    <gramGrp>...</gramGrp>
  </gramGrp>
  <gram type="root" lang="ara-Latn-x-1c">BKY</gram>
  <gram type="root" lang="ara">بَكَى</gram>
  <note type=""/>
</gramGrp>
  </form type="lemma">

```

Masdar

Arabic	<input type="text" value="بَكَى"/>	
Transliterated	<input type="text" value="bukā"/>	<input type="button" value="Delete"/>
Arabic	<input type="text" value="بَكَى"/>	
Transliterated	<input type="text" value="bukan"/>	<input type="button" value="Delete"/>



Dictionary Entries

TEI and XForms showing senses and synonyms

Senses

Syntactic note	+ min because of	Optional	Delete note
Syntactic note	+ acc	Optional	Delete note
Add syntactic note			
Translation	to cry	Language English	Delete this translation
Translation	to weep	Language English	Delete this translation
Add translation			
Translation	to bemoan	Language English	Delete this translation
Translation	to lament	Language English	Delete this translation
Translation	to mourn	Language English	Delete this translation
Add translation			
Theme/s	<ul style="list-style-type: none"> Chain armour Desert/wilderness Dignity/honour Dream/sleep Drought/failure of crops Emotion Enemy Fate/lot Fear 		
Add synonym Delete synonym			

```

</sense>
</sense>
<cit type="translation" lang="en">
  <lb>to cry</lb>
</cit>
<cit type="translation" lang="en">
  <lb>to weep</lb>
</cit>
<cit type="translation" lang="en">...</cit>
</sense>
<sense n="c03744b0beab584874d5a4579a0d3f4b">
  <cit type="translation" lang="en">
    <lb>to bemoan</lb>
  </cit>
  <cit type="translation" lang="en">
    <lb>to lament</lb>
  </cit>
  <cit type="translation" lang="en">
    <lb>to mourn</lb>
  </cit>
  <cit type="translation" lang="en">...</cit>
</sense>
</sense>
<usg type="hyper">emotion</usg>
<note type="optional">+ min because of</note>
<note type="optional">+ acc</note>
<note type="optional"/>
</sense>
  
```



Poems

TEI and XForms showing poem metadata

Alternate title

Transliterated

Arabic

Author and bibliographic data

Default author

Find author

Find attributed author

Find bib ref

Pages

Source

About the poem

Genre

Meter

Rhyme

Number of lines

Reason for quoting

- Another verse
- Biography
- Collection
- Content
- Diwan
- Expression
- Grammar
- Matal

```

<?xml header="1" encoding="UTF-8" version="1.0" standalone="no" >
<teiHeader>
  <fileDesc>
    <titleStm>
      <title lang="ara-Latn-x-ic">Mu'allaqa</title>
      <title lang="ara"/>
    </titleStm>
    <respStat>
      <resp author/>
    </respStat>
    <name>
      <ptr target="#d32440405a3f7ef5905543a7ec1110e">إمرء القيس (Imru al-Qays)</ptr>
    </name>
    </respStat>
    <respStat>
      <resp attributed/>
    </respStat>
    <name>...</name>
    </respStat>
  </fileDesc>
  <publicationStm>
    <distriButor>University of St Andrews</distriButor>
  </publicationStm>
  <sourceDesc>
    <listBibl>
      <bibl n="primary">
        <citeRange unit="pp">8-26</citeRange>
        <ptr target="#1">Dīwan Imru' al-Qays</ptr>
      </bibl>
      <bibl n="">
        <citeRange unit="pp"/>
        <ptr target=""/>
      </bibl>
    </listBibl>
  </sourceDesc>
  <fileDesc>
    <encodingDesc>
      <samplingDecl>
        <p>diwan</p>
      </samplingDecl>
    </encodingDesc>
    <profileDesc>
      <textClass>
        <keywords scheme="genre">
          <term>qasidah</term>
        </keywords>
        <keywords scheme="meter">
          <term>tawil</term>
        </keywords>
        <keywords scheme="rhyme">
          <term>i</term>
        </keywords>
        <keywords scheme="lines">
          <term>80</term>
        </keywords>
      </textClass>
    </profileDesc>
    <revisionDesc status="unread"/>
  </fileDesc>
</teiHeader>

```



Poems

TEI and XForms showing words in a line

Identify meanings of words in poem

Select line

 أَيْفَا نَيْكُ مِنْ نَكْرَى حَبِيبٍ وَمَنْزِلٍ بِسَجْطِ اللَّوَى بَيْنَ الدُّخُولِ فَحُؤْمَلٍ

Line number1

Hemistich1	نَيْكُ	مِنْ	نَكْرَى	حَبِيبٍ	وَمَنْزِلٍ	بِسَجْطِ	اللَّوَى	بَيْنَ	الدُّخُولِ	فَحُؤْمَلٍ
			DKR	HBB	NZL					
Hemistich2	فَحُؤْمَلٍ	الدُّخُولِ	بَيْنَ	اللَّوَى	بِسَجْطِ					
	HML	DIL	LWY	SQT						

Senses of words

Word

نَكْرَى

Prime reference

Root

DKR

Lemmata

Update senses

Update lemmata

Senses

Update syntactic notes

Syntactic note

New lemma

Translations of line

Translation

Stop, let us weep at the memory of a loved one and [her] dwelling at a place where the sands twist to an

English Delete

Add translation

Page number for line in primary source

Page

8

```

<?xml version="1.0" encoding="UTF-8" type="text/xml" >
<TEI xmlns="http://www.tei-c.org/ns/1.0" type="text" >
  <text >
    <div data-bbox="112 276 320 298" data-label="Section-Header" style="display:none" />
    <div data-bbox="112 311 156 325" data-label="Text" style="display:none" />
    <div data-bbox="112 326 430 348" data-label="Text" style="display:none" />
    <div data-bbox="112 363 180 378" data-label="Text" style="display:none" />
    <div data-bbox="225 398 464 528" data-label="Table" style="display:none" />
    <div data-bbox="112 550 187 566" data-label="Section-Header" style="display:none" />
    <div data-bbox="112 580 137 595" data-label="Text" style="display:none" />
    <div data-bbox="199 580 230 600" data-label="Text" style="display:none" />
    <div data-bbox="112 601 178 614" data-label="Text" style="display:none" />
    <div data-bbox="112 615 137 630" data-label="Text" style="display:none" />
    <div data-bbox="324 615 354 636" data-label="Text" style="display:none" />
    <div data-bbox="112 641 153 656" data-label="Text" style="display:none" />
    <div data-bbox="225 641 289 656" data-label="Text" style="display:none" />
    <div data-bbox="353 641 433 656" data-label="Text" style="display:none" />
    <div data-bbox="112 658 146 673" data-label="Text" style="display:none" />
    <div data-bbox="225 658 289 673" data-label="Text" style="display:none" />
    <div data-bbox="112 687 171 703" data-label="Text" style="display:none" />
    <div data-bbox="112 704 164 720" data-label="Text" style="display:none" />
    <div data-bbox="112 735 203 751" data-label="Section-Header" style="display:none" />
    <div data-bbox="112 769 467 846" data-label="Form" style="display:none" />
    <div data-bbox="112 859 293 876" data-label="Text" style="display:none" />
    <div data-bbox="112 890 137 904" data-label="Text" style="display:none" />
    <div data-bbox="200 891 211 908" data-label="Text" style="display:none" />
  </text>
  <source >
    <div data-bbox="472 218 893 960" data-label="Text" style="display:none" />
  </source>
</TEI>

```