

قضايا بحثية 2 - التكليف 5

Prepared by: **Abobakr Mohammed Baras**
ID: **253002656**

March 2026



هذه المهمة، سيقوم الباحث بتقييم رسالة ماجستير وتقييم طرق جمع البيانات
جب أن يشمل التقييم توضيح الطرق المستخدمة لجمع البيانات، بالإضافة إلى تفاصيل حول
فية تنفيذ هذه الطرق.

رسالة الماجستير

Quranic Arabic Wordnet: Design and Implementation

<https://search.mandumah.com/Record/1045509>

اعتمدت الرسالة على ثلاثة مصادر رئيسية للبيانات:

1. BAQ Corpus

○ يحتوي كلمات القرآن مع الجذور والتصنيفات النحوية.

2. القواميس العربية التقليدية

■ المعجم الوسيط

■ المعجم الغني

3. SALMA – ABCLexicon

○ قاعدة بيانات تحتوي جذور الكلمات العربية ومشتقاتها.

تم دمج هذه المصادر لبناء قاعدة بيانات تُستخدم لإنشاء **Quranic Arabic WordNet**.

- كلمات القرآن
- الجذور اللغوية
- التصنيف النحوي (POS)
- معلومات نحوية إضافية
- ترجمة إنجليزية للكلمات

كيفية جمع البيانات:

- استخدام نسخة BAQ corpus version 2.0
- استخراج كلمات الجزء الثلاثين (جزء عم) فقط
- باستخدام Python script

إحصائيات البيانات:

- 2308 كلمة قرآنية
- 749 اسم
- 437 فعل
- 532 جذر لغوي

تم تنفيذ عدة خطوات لمعالجة البيانات:

1. استخراج الكلمات القرآنية
 2. إزالة الكلمات الشائعة (Stop Words) مثل: (من، على، حتى)
 3. الاحتفاظ فقط بـ:
 - الأسماء
 - الأفعال
 - الصفات
- الهدف من ذلك:
- تقليل حجم البيانات
 - تحسين دقة التحليل الدلالي.

تم استخدام قواميس عربية تقليدية لاستخراج معاني الكلمات (Word Senses).

القواميس المستخدمة:

- المعجم الوسيط

- يحتوي حوالي 53,548 كلمة

- المعجم الغني

- يحتوي حوالي 29,803 كلمة

هذه القواميس توفر:

- تعريفات الكلمات

- معاني متعددة للكلمة الواحدة

تم تنفيذ عدة خطوات لاستخراج المعاني من القواميس:

1. Tokenization

- تقسيم النص باستخدام الفاصل |
- كل تعريف يُعتبر معنى مستقل للكلمة

2. تحديد نوع الكلمة (POS)

- استخراج الأفعال باستخدام نمط "فعل" والأسماء باستخدام نمط "اسم"

3. تخزين النتائج في قاعدة بيانات

- 157,738 معنى من المعجم الوسيط
- 80,528 معنى من المعجم الغني

قاعدة بيانات لغوية كبيرة تحتوي على:

- جذور الكلمات العربية
- جميع المشتقات الممكنة لهذه الجذور

تم استخدامها من أجل:

- استخراج مشتقات الجذور القرآنية
- زيادة التغطية اللغوية للكلمات.

خطوات جمع البيانات:

1. استخراج جذور الكلمات القرآنية من BAQ corpus.
2. استخدام هذه الجذور للبحث في SALMA Lexicon.
3. استخراج كل المشتقات الممكنة لهذه الجذور.
4. إزالة اللواحق والبادئات (Clitics) يدويًا مثل:

- و
- ال
- ضمائر الملكية.

عدد الكلمات المشتقة الناتجة:

حوالي 80,000 كلمة.

1. تنوع مصادر البيانات

- Corpus
- قواميس
- Lexicon

2. الاعتماد على مصادر لغوية موثوقة

3. استخدام طرق آلية في استخراج البيانات

- Python
- Regex
- Tokenization

4. بناء قاعدة بيانات منظمة للمعالجة.

الاعتماد على جزء واحد فقط من القرآن (جزء عم)
مما يقلل من شمولية البيانات.

بعض العمليات تمت يدويًا
مثل:

- إزالة clitics
- تحديد POS لبعض الكلمات

احتمال وجود أخطاء في المعالجة التلقائية للنصوص العربية.

القواميس المستخدمة قد تحتوي تعاريف قديمة أو غير سياقية.

بشكل عام:

- استخدمت الرسالة منهجية مناسبة لجمع البيانات.
- تم الاعتماد على مصادر لغوية قوية.
- تم تطبيق معالجة لغوية حاسوبية باستخدام أدوات برمجية.

لكن يمكن تحسين العمل من خلال:

- استخدام كامل نص القرآن
- تقليل العمليات اليدوية
- استخدام أدوات أكثر تطورًا لمعالجة اللغة العربية.



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

