

أهمية المصادر المفتوحة في المعالجة الآلية للغة العربية

مقدمة

إن الحاجة الملحة للتطبيقات اللغوية لم تعد أمرا خفيا، لا سيما في مجال المعلومات وتكنولوجياها، وصار الحديث عن المعالجة الآلية للغة الطبيعية أمرا مهما في البرامج الحديثة من بواحث وقواميس وترجمة وغيرها من الحاجات اليومية للمستخدم العادي.

وقد انكبت كثير من مراكز البحث والتطوير على دراسة اللغة العربية واستخراج أسرارها من أجل معالجتها معالجة آلية، لكن هذه الجهود والنتائج قد تلقى مصيرين لا ثالث لهما، إما أن تبقى حبيسة الجامعات أو المراكز، أو حكرًا على المؤسسات الضخمة التي تمول هذه المشاريع، مما يجعل تكلفة إدماجها في الحلول البرمجية البسيطة أمرا غير ذي جدوى اقتصادية، فنجد قواميس ومحركات بحث لا تأخذ بعين الاعتبار إمكانيات العربية وخصوصياتها.

و من هنا لاحظنا كغيرنا الجهود التي يبذلها الباحثون مرارا وتكرارا للحصول على محلل صرفي أو نحوي أو مدقق من أجل المضي قدما في أبحاث المعالجة الآلية، فيبذلون جهدا جهيدا لإعادة اختراع عجلة، عجلة قد لا تكون مستديرة تماما.

و خلال بحثنا في أطروحة الدكتوراة "المعالجة المعلوماتية للقرآن وعلوم القرآن: مقارنة معنوية أنطولوجية"، وعند مرحلة إنجاز وحدة البحث في النص القرآني، واجهنا مشكل البحث الصرفي، الذي ليس من صميم بحثنا، فإذا بحثنا في هذا الموضوع نجد الكثير من برامج البحث الصرفي التجارية المكلفة، لا يمكن إدراجها حتى بعد شرائها بأثمان باهظة ضمن بحوثنا نظرا لقيود الحقوق التجارية والملكية الفكرية. أي أننا باختصار وقعنا في فخ العجلة غير المستديرة.

لذا فإن فلسفة المصدر المفتوح تعطي حلا مناسبًا يمكن من خلاله تقاسم المعارف والجهود، وعدم الاختباء وراء الحقوق الملكية التي تعطل أحيانا تطور الأبحاث وتظافر الجهود.

دراسة تجارب عملية للمدققات الإملائية

بعد حصول دعم اللغة العربية في المدققين الإملائين Aspell و Hunspell انبرى الأخ محمد سمير¹ من مصر فاستعمل - للاختبار - في أول وهلة القرآن الكريم كقاموس مرجعي لهذا المدقق الإملائي

¹ -<http://home.foolab.org/cgi-bin/viewcvs.cgi/projects/arspell>

ثم بعدما لاحظ عدم صلاحية هذا الحل توجه إلى قاموس QAMUS LLC² لصاحبه Tim Buckwalter واعتمده قاعدة لمشروع المدقق الإملائي³ للغة العربية إلا أنه تبين له بعد ذلك عدم نجاعة هذا الاختيار.

كما استعمل قاموس QAMUS LLC أيضا في مشروع Ivrix⁴ للمطور الإسرائيلي Dan Kenigsberg ثم لحق بهما فريق غوغل⁵ Google الذي ضم بين ثناياه قاموس Tim Buckwalter affixe file نفسه فأصدروا إصداراً يحوي لأول مرة ملف زوائد. نلاحظ أن المشاريع كلها لجأت لقاموس Tim Buckwalter بسبب غياب قواميس حرة أخرى بديلة للغة العربية.

و قد لجأنا في مبادرتنا إلى الاعتماد على خوارزمية هانسبل Hunspell وقاموس Tim Buckwalter في البداية وملف زوائد لتجربة أقصى حدود هانسبال.

المقارنة بين المدققات الإملائية الموجودة

الدولي	البرنامج
محمد الزبير	المطور
بايثون/ سي	لغة البرمجة
لينكس / متعدد الأنظمة	نظام التشغيل
BSD	الترخيص
عيون العرب Arabeyes	الجهة الداعمة
Arabeyes.org	الموقع
http://www.arabeyes.org/project.php?proj=Duali	
قاموس Tim BulkWalter	القاموس
دعم لغة الضاد لأول مرة.	الوصف
مدقق عربي منذ البداية.	مزايا
توقف التطوير به، واجهة نصية..	مساوئ

² -<http://www.qamus.org/>

³ -<http://home.foolab.org/cgi-bin/viewcvcs.cgi/projects/arspell>

⁴ -<http://www.ivrix.org.il/projects/arabic/>

⁵ -<http://sourceforge.net/projects/arabic-spell/>

البرنامج	بغداد
المطور	محمد سمير
لغة البرمجة	C
نظام التشغيل	لينكس (All POSIX (Linux/BSD/UNIX-like OSes)
الترخيص	BSD
الجهة الداعمة	عيون العرب Arabeyes.org
الموقع	http://home.foolab.org/cgi-bin/viewcvs.cgi/projects/baghdad
القاموس	قاموس Tim BulkWalter
الوصف	(fork duali) تحويل لغة دُولي من لغة برمجة بايتون إلى C++
مزايا	
مساوئ	متوقف التطوير...

البرنامج	Ivrix's spell-ar
المطور	Kevin Atkinson
لغة البرمجة	/ C
نظام التشغيل	لينكس / متعدد الأنظمة
الترخيص	GNU General Public licence / BSD
الجهة الداعمة	http://www.ivrix.org.il
الموقع	http://www.ivrix.org.il/projects/arabic
القاموس	قاموس Tim BulkWalter
الوصف	برنامج موجه أصلا للغة العبرية. اختبار الدعم العربي في Aspell بتوظيف
مزايا	قاموس Tim BulkWalter The word list was repackaged in Aspell's format
مساوئ	20Mo، غياب ملف الزوائد، أخطاء كثيرة.

البرنامج	Google-arspell
المطور	Google Team
لغة البرمجة	C
نظام التشغيل	– متعدد الأنظمة
الترخيص	GNU General Public License (GPL), GNU Library or (Lesser General Public License (LGPL
الجهة الداعمة	.Google inc
الموقع	https://sourceforge.net/projects/arabic-spell
القاموس	قاموس Tim BulkWalter
الوصف	قائمة كلمات وملف زوائد.
مزايا	
مساوئ	ثقيل، مكنز، أخطاء كثيرة.

البرنامج	Hunspell
المطور	نامث لاسي، وطه زروقي للمزايا العربية
لغة البرمجة	C
نظام التشغيل	متعدد الأنظمة
الترخيص	/LGPL/MPL
الجهة الداعمة	OpenOffice
الموقع	Hunspell.sourceforge.net Ayaspell.blogspot.com
القاموس	قاموس Ayaspell
الوصف	إعداد قاموس لبرنامج هانسبال إضافة المزايا العربية لهانسبال
مزايا	خصائص إتصاقية + خصائص اشتقاقية شبه – توليدية. اعتماد المعاجم العربية، اعتماد قواعد اللغة الصرفية والنحوية.
مساوئ	في طور الإنشاء (:

المراجع

1. موقع المشروع <http://ayaspell.blogspot.com>

3. Ispell المدقق الإملائي <http://www.gnu.org/software/ispell/ispell.html>
4. Aspell المدقق الإملائي <http://aspell.sourceforge.net/>.
5. Hunspell . <http://hunspell.sourceforge.net>
6. Duali : <http://www.arabeyes.org/project.php?proj=Duali>
7. Mohamede Sameer : araspell project, <http://www.foolab.org/projects/arspell>
8. Ivirix Project : <http://www.ivrix.org.il/projects/arabic/>
9. Google team for arabic-spell : <http://sourceforge.net/projects/arabic-spell/>
10. Tim Buckwalter: aramorph project : <http://www.qamus.org>