

المصحح الآلي "سراج"

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

جامعة الملك فهد للبترول والمعادن

مادة تعريبيه الحاسبه ICS 484

السنة الدراسيه 2006\2007

إعداد

زياد احمد العيسى

212453

مواد محاضريه

أستاذ المادة

حسني المحتسب

2006-12-12

ملخص المحاضرة

نبذة عن صخر:-

تأسست شركة صخر في عام 1982 كإحدى شركات مجموعة العالمية، وطورت العديد من التقنيات الفائقة التي تركت بصمة قوية في مجال صناعة تقنية المعلومات العربية.

صخر تسعى الى تطوير الجيل الجديد من تقنيات المعالجة الطبيعية للغة العربية (NLP) في الوقت الذي ابتعد فيه مطورو البرامج العالميون عن بدء أي أبحاث جادة أو تطوير تقنيات لدعم لغة صعبة مثل العربية .

في الوقت نفسه لم تكن لدى مطوري البرامج المحليين الكفاءة لخوض المنافسة التقنية ولا الموارد المالية اللازمة للبدء في خوض مثل هذه التجربة الطموحة .

كانت صخر هي الوحيدة التي لديها الخبرة والموارد، فضلاً عن تمتّعها بالعزيمة القوية والبصيرة النافذة لتتفقد سنوات من البحث المكثف في اللغة العربية واستخدام الحاسب بها .

لقد كان تقليد صخر للتحديث المستمر هو السر وراء النجاح في إنتاج أول حاسب منزلي باللغة العربية في العالم، حيث كان برهاناً على الحاجة الملحة لتوفير برامج باللغة العربية .

وللمحافظة على هذا السبق حتى الآن، استطاعت صخر أن تصبح الشركة الرائدة في مجال تطوير الحلول والمنتجات باللغة العربية، وتغلبت على سيادة اللغة الإنجليزية في مجال صناعة المعلومات العربية.

قامت صخر، في الوقت نفسه، بتصحيح الاعتقاد الخاطئ بأن الحلول المطورة في الغرب يمكن أن تُكيف لتلبي احتياجات المستخدم العربي.

وفي سبيل توفير حلول ذكية ومتطورة إلى أقصى درجة للمستخدم العربي، واجهت صخر العديد من التحديات الصعبة التي أحبطت العديد من المطورين على المستويين المحلي والعالمي.

من بين هذه التحديات انعدام الدعم المالي من قبل الحكومات، مما دفع بصخر إلى البدء في إجراء أبحاث ومشاريع تطوير استغرقت أكثر من مليوني ساعة عمل بشرية، ممولة بالكامل بواسطة الاستثمار

الخاص. انطلاقاً من إيماننا العميق بقيمة اللغة العربية للعملاء من أفراد وشركات وحكومات.

صخر هي الشركة الرائدة في تقنيات اللغة العربية في مجالات الكلام والترجمة الآلية والبحث والقارئ الآلي والتعليم وإدارة المعلومات. وبمقرها في مصر يعمل لدى الشركة حوالي 300 من خبراء تقنية المعلومات، فضلاً عن مكاتبها في جميع أنحاء الشرق الأوسط

المصحح الآلي "سراج":-

يعتبر "سراج" أحد تلك البرامج العربية التي تسعى صخر لتطويرها مجال في تطبيق منهجيات البحث والتصحيح في النص داخل قواعد بيانات النص العربي. يقوم سراج بعمليات تحليل للنصوص مثل تصحيح النص و استخلاص كلماته الأساسية و تصنيفه و تلخيصه و استخراج الأعلام.

وهذه المكونات الأربعة هي:

مصحح صخر

- يصحح كل من النص العربي والنص الإنجليزي على حد سواء
- يصحح الأخطاء العربية الشائعة
- يصحح الأخطاء النحوية العربية
- يمكنه العمل كمصحح إملائي دقيق دون الحاجة للتدخل البشري
- يمكنه ربط النص بأحد الطرق المنهجية للكتابة العربية
- وضع خاص لتوحيد الأسماء والتصحيح

يقوم المصحح الآلي من صخر باكتشاف وتصحيح الأخطاء الإملائية العربية والأخطاء العربية الشائعة إضافة إلى الأخطاء النحوية.

يجب القيام بهذا التصحيح قبل القيام بعمليات تحليل النصوص الأخرى. فيجب أن يكون النص المدخل بهذه العمليات صحيحاً قبل أن يتم استخلاص كلماته الأساسية أو تصنيفه أو تلخيصه أو ترجمته أو البحث في محتواه أو تنفيذ أي عمليات أخرى عليه.

مستخلص الكلمات المفتاحية

- التقسيم الآلي للجمل
- استخلاص الجمل والكلمات المفتاحية

- استخلاص أسماء الأعلام بما يشمل أسماء الأشخاص والأماكن والمؤسسات المعروفة إلى جانب استخلاص العبارات المشهورة... الخ

تقوم أداة استخلاص الكلمات الأساسية بتحليل أي مستند عربي، والتعرف على عبارات النص وعناصر البيانات الرئيسية تلقائياً.

وهي تساعد الشركات على تطوير فرص العمل الرئيسية عن طريق تحديد المعلومات الهامة الموجودة في البيانات الضخمة، ثم استخلاص التفاصيل الأكثر أهمية بذكاء ليتم استخدامها بالمؤسسة.

تساعد الكلمات الأساسية على تصنيف المستند في شجرة الموضوعات المحددة من قبل المستخدم لتسهيل استعراض المعلومات والوصول إليها. كما تساعد هذه الأداة كذلك على ربط المستندات ببعضها بطريقة ديناميكية، وتقليل الوقت المطلوب للبحث في محتواها .

يوفر لك محرك أداة استخلاص الكلمات الأساسية العربية من صخر إمكانية الوصول السريع إلى المزايا والعبارات الفردية لاستخدامها في إنشاء ملخصات النصوص وتصنيف المعلومات وأنشطة تحليل النصوص الأخرى.

المصنف

- ينظم محرك مصنف صخر المعلومات القيمة بدقة وكفاءة على شكل شجرة مواضيع منطقية أو على شكل تصنيف لمحتوى المشروع .
- يوظف محرك التصنيف مصحح صخر ليقوم بالتصحيح التلقائي للنص العربي المدخل من الأخطاء العربية الشائعة
- تشمل شجرة التصنيف في هذه العملية ما يصل إلى 20 تصنيف عام رئيسي بما في ذلك الفنون والعلوم والديانات والسياسات... الخ .
- تحتوي شجرة التصنيفات أيضاً على 100 تصنيف فرعي مندرج تحت التصنيفات الرئيسية وكل هذا من أجل ضمان الدقة.

ملخص صخر

- يتعرف ملخص صخر العربي على أكثر الجمل ذات الصلة داخل النص ويعرضهم على شكل ملخص للنص .
- يقلل بالفعل الوقت الذي نحتاجه لقراءة المعلومات ومعالجتها يدويًا .
- يوفر للشركات ملخصًا قصيرًا للنص لكل بند باستخدام قائمة تضم الجمل المفتاحية مرتبة حسب أهميتها في النص ويوفر إمكانية اختيار مستوى معين من التلخيص .
- يقوم الملخص الآلي باستخلاص الجمل المفتاحية الهامة من الوثيقة بناءً على قواعد إحصائية ولغوية (على سبيل المثال: نمط الكتابة، وضع الجملة في الوثيقة، الكلمات المفتاحية المتضمنة، أنواع الكلمات المفتاحية... إلخ)
- يمكن للعميل أن يختار الملخص الذي ينتج عن النص وفقًا لنسبة مئوية من المستند المدخل، أو وفقًا لعدد ثابت من الجمل أو حجم النص .
- يوظف محرك التلخيص مصحح صخر ليصحح النص العربي المدخل تلقائيًا من الأخطاء العربية الشائعة، كما يوظف مستخلص الكلمات المفتاحية ليتعرف على قائمة بالكلمات المفتاحية الهامة وهذا للتعرف بدقة على الجمل الهامة.

يقوم محرك الملخص العربي من صخر بتحديد أكثر الجمل أهمية في النص وعرضها كملخص نصي قصير. يسهل الملخص إجراء مسح للجمل الهامة بالمستندات، مما يوفر الوقت المطلوب لقراءة البنود ومعالجتها يدويًا.

يوفر هذا المحرك للشركات إمكانية التلخيص النصي لجميع البنود باستخدام قائمة كلمات أساسية مرتبة حسب الأهمية، مع إمكانية تحديد مستوى معين للتلخيص.

يقوم محرك التلخيص بتوظيف المصحح الآلي من صخر لتصحيح النص العربي المدخل تلقائيًا وتنقيته من الأخطاء العربية الشائعة، كما يوظف مستخلص الكلمات الأساسية لتعريف قائمة كلمات أساسية مرتبة حسب الأهمية لتحديد الجمل الهامة.

متطلبات البرنامج

- خادم بمعالجين XEON

- نوافذ مايكروسوفت 2000 أو 2003
- ذاكرة مؤقتة 1 جيجا بايت
- 500ميغا بايت مساحة خالية على القرص الصلب

المراجع المستخدمة

- موقع شركة صخر www.sakhr.com
- النسخة التجريبي سراج

المشاكل التي واجهتني

لا توجد مشاكل كثير باستثناء كثرة المعلومات عن البرنامج وكيفية شرحها ترتيبها بشكل واضح وممتع.

الفوائد

استفدت كثيراً من تجربتي هذه ومن هذه الفوائد:-

- معرفة معلومات وفيرة عن شركة صخر وعن منتجاتها
- طرق جميله ورائعة في مجال المصحح الآلي

بالنسبة لاستفادتي من تقديم الحاضرة فأهم ماتعلمته هو اكتساب الثقة بشكل أكبر كونها ليست المرة الأولى التي أقدم بها محاضره، أيضاً محاولة الاستفادة من الأخطاء التي ارتكبتها لتطوير ذاتي مجدداً

الأسئلة

أجب بصح أو خطأ:-

[صح] يراعي مصحح الأخطاء سياق الجملة عند إعطائه المقترحات للأخطاء الإملائية.
[خطأ] لا يتعامل المصحح الآلي مع التشكيل بالعربي

أختر الإجابة الصحيحة:-

1) ماهي مكونات المصحح الآلي سراج:

- أ - مصحح صخر ، مستخلص الكلمات المفتاحية ، المصنف ، ملخص صخر
- ب- مصحح صخر ، مستخلص الكلمات المفتاحية ، المصنف
- ج- مصحح صخر ، المشكل الآلي ، القارئ الآلي ملخص صخر
- د - مصحح صخر ، المصنف

2) تعتبر شركة صخر من الشركات الرائدة في :-

- أ - برامج الألعاب
- ب- برامج تقنيات المعالجة الطبيعية للغة العربية
- ج- نظم التشغيل
- د- تعريب البرامج الغربية