

مشروع عملي نظم استرجاع المعلومات

المطلوب:

بناء نظام استرجاع معلومات information retrieval system قادر على إضافة وثائق نصية في مجال معين والقدرة على الاستعلام باللغة الطبيعية نصياً عن معلومات ضمن هذه الملفات
يجب على كل مجموعة بناء نظام تحقيق ما يلي:

1. بناء وتحقيق نظام استرجاع معلومات يحتوي على الأقل:

- معالجة نصية للملفات - القدرة على تجذير الكلمات باستخدام أحد خوارزميات التجذير المعروفة.
- معالجة نصية للملفات - القدرة على الوصول للكلمات الشاذة (كالأفعال غير النظامية)
- معالجة نصية للملفات - القدرة على الوصول إلى التعابير التي تملك أشكال مختلفة (كالتواريخ)
- القدرة على اقتراح تصحيح للكلمات في حال إدخال استعلام بصيغة لغوية خاطئة
- القدرة على الوصول للأسماء باختلاف طرق التمثيل الكتابي لها.
- بناء فهرس مناسبة للمحتوى المعالج في المراحل السابقة.
- بناء تابع مطابقة بين الاستعلام والمحتوى السابق
- search UI

2. إضافة واحدة من الميزات التالية للنظام: (من 7 إلى 10 علامات)

- Query refinement (query autocompletion, query suggestions)
- Use advanced word embedding models
- Crawling
- Document clustering
- Use learning to rank models
- Personalization
- Use link analysis
- Topic detection
- Multimedia retrieval
- Multi lingual retrieval system
- يمكن اقتراح ميزة مختلفة على أن يتم الموافقة عليها من قبل أستاذ العملي أو الدكتور.

3. كتابة تقرير مفصل باللغة العربية وعرض يحتوي على شرح ما يلي:

- توصيف الداتا سيت المستخدمة (يجب بناء النظام على 2 datasets) المقترحات وليس إحداها
- توصيف خطوات المشروع ولكل service ضمن النظام
- توصيف لبنية النظام system architecture يوضح فيه البنية لل services وكيفية التواصل بينها
- النتائج والتقييم على ال testing data ضمن الداتا سيت المرفقة, المطلوب حساب:
Precision@10 - precision - recall - Mean Average Precision (MAP) - Mean Reciprocal Rank (MRR)
- إعادة حساب التقييم بدون ومع الميزة الإضافية للنظام مع كتابة تحليل للنتيجة
- تقسيم العمل بين أعضاء المجموعة
- المصادر

4. إحضار نسخة تنفيذية للنظام ضمن المقابلة

5. رابط GitHub للكوود البرمجي الخاص بالمشروع مع readme واضح لبنية الكود.

6. تسليم التقرير بشكل مفصل باللغة العربية (عليه جزء من العلامة)

يرفق بنص المشروع 2 datasets مع أمثلة للاستعلامات عليهما والنتائج الأكثر ملائمة لهذه الاستعلامات، سيتم تقييم المشاريع بناء على ال Testing Collection المرفقة بالمشروع.

ملاحظات تنظيمية:

- لا يوجد أي قيود على اللغات أو المنصات الممكن العمل بهل والاعتماد عليها طالما أنها تعطي نتائج صحيحة ومنطقية.
- يمكن أن تحتوي المجموعة على 3-4 أفراد (في حال كان العدد 5 المطلوب تحقيق ميزتين إضافيات من الطلب الثاني)
- يمكن استبدال واحدة من ال datasets المقترحة بأخرى خارجية

التسليم:

موعد التسليم النهائي 05-06-2022

المقابلات ستتم على مرحلتين:

المرحلة الأولى يتم فيها متابعة التطور المرحلي الخاص بكل مجموعة ضمن محاضرات العملي (مرة واحدة لكل مجموعة)

المرحلة الثانية هي المقابلات النهائية يوم 09-06-2022

مدّرس النظري: د.م. أبي صندوق

مدّسة العملي: م. لين قويدر