تود شركة Video Track الاعتماد على مخزن الأفلام الموجود على بعض مواقع الأفلام والمرفق بنصوص ترجمة باللغة الإنكليزية (Subtitle) من إجراء عملية تسهل البحث الدلالي في أجزاء هذه الأفلام اعتماداً على نصوص الترجمة المتوفرة ولحظات الظهور المرافقة لها.

المطلوب بناء نظام يوفَر الوظائف التالية:

1. تمكين التصفح عبر مجموعة من المواقع يحدَدها مدير النظام و جمع ملفات subtitles المتوفرة و المرتبطة بوصلات الافلام المتوفرة في الصفحة، واستخراج ملفات نصوص الترجمة الموجودة ضمن مكان محدًد في هيكلية الأفلام المعدَة للعرض.
2. بناء مخزن بيانات شخصي من ملفات RDF تُولد آلياً، يرتبط بكل جزء من ملفات الفيديو والتي تحددها النقاط الزمنية لإظهار نصوص الترجمة بعد استخراج ثلاثيات (Subject, predicate , object) من نص الترجمة وإضافة الوسم الزمني كإحدى الخصائص، وأية خصائص أخرى تراها ضرورية لتحقيق الغرض.

**ملاحظة** : يمكنك استخدام آلية بسيطة فقط لاستخراج معلومات الجملة الأولى من كل سطر ترجمة أو الاستعانة هنا بمكتبات معالجة اللغات الطبيعية <http://nlp.stanford.edu/software/lex-parser.shtml#Sample> ، <http://opennlp.apache.org>

1. قم باستخدام أحد أدوات تقييم ملفات RDF للتحقَق من صحة صياغة الملفات المولدة

) <http://www.w3.org/RDF/Validator/>(

1. ناقش إمكانية تحسين جودة محتوى الملفات المولدة يدوياً بعد توليدها و واجهة المستخدم المطلوبة لهذا الغرض .
2. تمكين عملية الاسترجاع بالاعتماد على المعلومات التي تم توفيرها من خلال ملفات RDF المنشأة باستخدام استعلامات بلغة Sparql(يمكن استخدام مكتبة vds.rdf.query أو http://graphite.ecs.soton.ac.uk/sparqllib لهذا الغرض)!
3. قم باختبار لربط الموارد التي قمت بإنشائها مع DBPedia أو Dublin core أو FOAF لمحاولة مقابلة المفاهيم المستخدمة في المشروع مع المفاهيم المعتمدة في هذه الصفوف من البيانات .
4. بناء الواجهات البرمجية المناسبة لعملية البحث من خلال الكلمات المفتاحية أو من خلال exploratory Navigation,

**مراحل التنفيذ:**

**المرحلة الأولى :**

تتضمن عملية:

* 1. بناء الوكيل البرمجي الذي يقوم بجمع البيانات.
  2. اختيار الهيكلية وفضاء الأسماء الذي سيستخدم ضمن ملفات RDF لتمثيل مجموعات البيانات التي ستجمع.
  3. بناء التطبيق المسؤول عن صياغة و تخزين صفوف بيانات RDF ( يمكن اختيار الصيغة الأنسب للتخزين)
  4. التحقق من صفوف مجموعات البيانات في المخزن الشخصي الذي تم إنشاؤه.

**المرحلة الثانية**

1. اقتراح آليات يدوية لتحسين صفوف البيانات وبناء نموذج أولي للواجهات الممكن استخدامها لهذا الغرض!
2. تحقيق آلية للاستعلام عن المعلومات في المخزن الشخصي الذي تم إنشاؤه لصفوف البيانات!
3. ربط المخزن الشخصي الذي تم إنشاؤه بأحد المخازن العامة مثل DBPedia
4. بناء واجهات برمجية لاستثمار صفوف البيانات الناتجة

**الوثائق المطلوب تسليمها:**

* مخطًطات use case , analysis class diagram and activity diagram
* وثائق ERD أو XML schema (بحسب المقارية المستخدمة( إضافة إلى مخطًط معمارية النظام ومخطط Class diagram
* Name space schema المستخدمة ضمن ملفات RDF.
* تطبيق برمجي عامل يتضمًن كامل الوظائف المحددة (يترك للطالب حرية اختيار بيئة العمل ولغة التطوير المستخدمة)
* تقرير تفصيلي يتناول عرض للأدوات والتقنيات التي تم استخدامها لتحقق النظام المطلوب.

**المخرجات العلمية:**

* فهم آلية عمل برمجيات الوكلاء البرمجيين.
* فهم آلية بناء ملفات RDF كأحد طرق تمثيل البيانات الوصفية (Meta Data) و طريقة التعامل معها.
* استيعاب مفهوم البيانات المترابطة (linked data) و المبادئ الأربعة الأساسية التي تعتمدها.
* استكشاف مفاهيم تمثيل البيانات الوصفية الخاصة بالأغراض متعدَدة الوسائط.

د.سامي خيمي