## RiTa library:

مكتبة برمجية تحتوي على مجموعة من الأدوات البرمجية , هي مكتبة مفتوحة المصدر أي يمكن الحصول عليها بشكل مجاني , يمكن استخدامها مع العديد من البيئات البرمجية , تم تصميمها لدعم إنشاء أعمال جديدة تحتاج إلى معالجات على اللغة المكتوبة و النصوص , توفر الأدوات اللازمة للكتاب أو الفنيين أو المبرمجين الذين يتعاملون مع اللغات الطبيعية , مصممة بطريقة لكي تكون بسيطة الاستخدام , من ضمن الميزات التي تملكها :

* القواعد ( Grammar )
* Markov-based generation to text-mining
* POS ( part-of-speech )
* الصوتيات (phonemes )
* و العديد من الميزات الأخرى

جميع التوابع الموجودة ضمن هذه المكتبة مجربة و لا تحتاج إلى أي عمليات تدريب على بيانات محددة , كما يمكن أيضا دمج مكتبة RiTa و مكاملتها مع معاجم خاصة أو مع قاعدة بيانات ( WordNet).

مكتبة Rita لا تستخدم تقنيات التعلم الالي (machine learning techniques ) و إنما هي مجموعة من الحلول الاستدلالية لمجموعة محدودة من مهام اللغة الطبيعية.

يمكن استخدام هذه المكتبة و تنفيذ توابعها عن طريق لغات البرمج التالية: Java , JavaScript و و تعمل على العديد من البيئات البرمجية مثل:

Android, Processing, Node JS.

## شرح عن بعض ميزات المكتبة:

* **الصوتيات ( phonemes ):**

إحدى ميزات تحليل النصوص و هي عبارة عن وحدة صوت واحدة لها معنى في أي لغة , كل وحدة صوتية تمثل صوت مختلف , و يوجد تقريباً 40 وحدة صوتية يمكن تميزها.

مثال: الحرفين ch يوجد لديها ثلاث تهجيات مختلفة بثلاث صوتيات.

مكتبة Rita تستخدم Arpabet من أجل تمثيل الصوتيات.

Arpabet: هي عبارة عن أكواد نسخ صوتية تم تطويرها من قبل Advanced Research Project Agency (ARPA), يتم فيها تمثيل كل صوت لغوي إنكليزي بتسلسل مميز من محارف ASCII.

* **Stresses:**

يشير الضغط هنا إلى درجة التشديد على الصوت أو المقطع في الكلام , مما يؤدي إلى ارتفاع الصوت بشكل نسبي.

* **POS(Part-Of-Speech):**

تعبر عن الخصائص النحوية لكل كلمة حسب موقعها من الجملة و أجزاء اللغة الإنكليزية شائعة الاستخدام هي:

Noun, verb, adjective, adverb, pronoun, preposition, conjunction, article, etc.

* **التجذير ( Stemming ):**

و تعني إعادة الكلمة إلى قاعدتها الأساسية أو جذرها الأساسي , و عملية التجذير هي عملية هامة من أجل و مفيدة من أجل القيام باي نوع من أنواع معالجة النصوص.

* **Tokenizing:**

هي عملية تقطيع النص إلى أجزاء أصغر تسمى الرموز المميزة, يوجد العديد من الطرق لتقطيع النص, الطريقة التي تستخدمها مكتبة Rita تدعى Penn Treebank convention.

يمكن الحصول على الميزات السابقة عن طريق استخدام التوابع التالية من مكتبة Rita:

* RiTa.getPhonemes
* RiTa.getStresses
* RiTa.getPosTags
* RiTa.stem
* RiTa.tokenize