

التلخيص الآلي للنصوص العربية باستخدام الأنطولوجيا

Automatic Summarization of Arabic Text Using Ontology

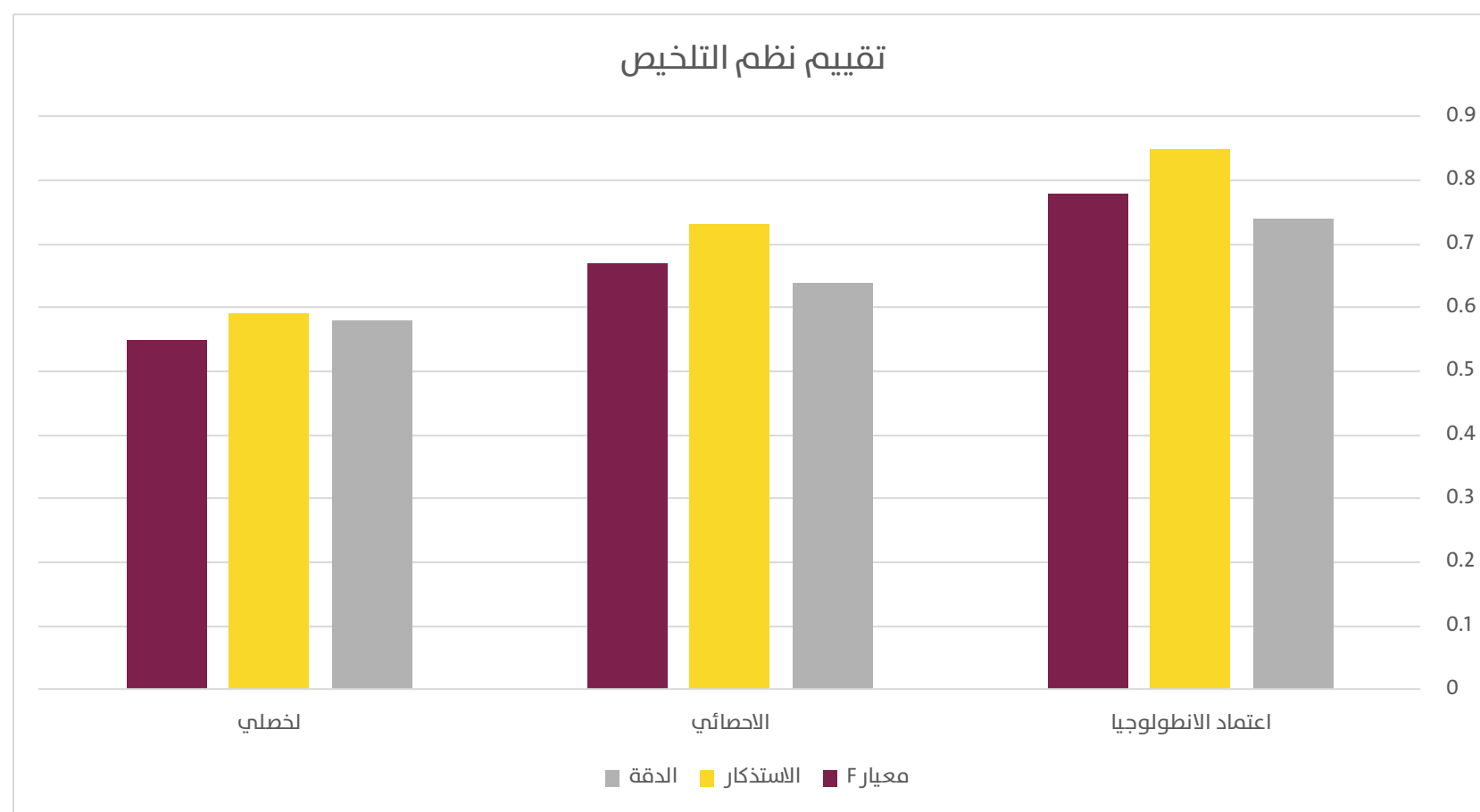
المهندسة سندس محمد سعيد الحلبي

الدكتور المشرف : الأستاذ الدكتور محمد مازن المحاييري

النتائج والمناقشة

استُخدمت مجموعة الملخصات العربية من (Essex (EASC (El-Haj et al., 2010) لاختبار نظم التلخيص المقترحة وتقييمها.

حُسبت مقاييس الدقة والاستذكار والمعامل F لملخصات كل من نظام التلخيص الإحصائي، ونظام التلخيص الدلالي، ونظام التلخيص المشهور لخصلي Lakhasly والذي يعد أحد أبرز المواقع التي تقوم بتلخيص الفقرات والنصوص العربية على الإنترنت Online وأكثرها استخداماً. وقد تبين أن متوسط الدقة والاستذكار ومعامل F في التلخيص المعتمد على الأنطولوجيا هي الأفضل مقارنة مع انظامين المقارن معهما وهي نسب جيدة مقارنة مع أنظمة التلخيص العربية عموماً. كانت متوسط الدقة والاستذكار و معامل F في التلخيص باعتماد الأنطولوجيا هي: 0.74، 0.85، 0.78، على الترتيب. وفي التلخيص الإحصائي: 0.64، 0.73، 0.67، على الترتيب. وفي نظام تلخيص لخصلي هي: 0.58، 0.59، 0.55، على الترتيب. والشكل التالي يوضح مقارنة نتائج تقييم التلخيص باعتماد الأنطولوجيا مع التلخيص الإحصائي وتلخيص لخصلي.



المراجع

- Nazari N., Mahdavi M., "A survey on Automatic Text Summarization." Journal of AI and Data Mining, 2019: 121-135
- Imam I, Hamouda A. Khalek HAA., "An ontology-based summarization system for arabic documents (ossad)." International Journal of Computer Application 74(17):38-43, 2013.
- Lloret E., Ferrández O. Muñoz R. Palomar M., "A Text Summarization Approach Under the Influence of Textual Entailment." 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008). 2008. 22-31.
- El-Haj M., Kruschwitz U. Fox C., "Multi-document arabic text summarisation." Computer science and electronic engineering conference (CEEC), 2011 3rd, IEEE Xplore, 2011b: 40-44.
- El-Haj M., Kruschwitz Udo, Fox Ch. "Using mechanical turk to create a corpus of arabic summaries." 2010.

الملخص

تلخيص النصوص هو عملية استخلاص أو تجميع المعلومات المهمة من النصوص الأصلية وعرضها بشكل ملخص. يوجد الكثير من الأبحاث والدراسات والأنظمة الخاصة بالتلخيص الآلي للنصوص باللغات الأوروبية المتعددة وغيرها من اللغات العالمية. إلا أن الأبحاث والدراسات في مجال تلخيص النصوص العربية ضئيلة وتنمو ببطء (El-Haj et al., 2011) واتبعت في معظمها الأسلوب الاستخراجي واستخدام تقنيات إحصائية وابتعدت عن الأسلوب التجريدي واستخدام التقنيات الدلالية استناداً إلى مصادر معرفية مثل الأنطولوجيا.

لذا اقترح في هذا البحث نظام تلخيص آلي للنصوص العربية باعتماد الأنطولوجيا. وقد بينت النتائج أن متوسطات الدقة والاستذكار ومعامل F في التلخيص باعتماد الأنطولوجيا المقترح هي نسب جيدة مقارنة مع نسب أنظمة التلخيص الآلي للنصوص العربية المقارن معها، ومع أنظمة التلخيص عموماً.

القسم النظري

صُنفت أنظمة التلخيص (Lloret et al., 2008). إلى أحادية المستند أو متعدد المستندات، عامة أو مقادة بالاستعلام. استخراجية أو تجريدية، وقد استخدمت تقنيات عدة في تلخيص النصوص التي اعتمدت في عدة مشاريع في تلخيص نصوص في اللغات الأوروبية عموماً، وفي اللغة العربية خصوصاً كما في نظام OSSAD (Imam et al., 2013)، لأنها استخدمت أنطولوجيات مترجمة تفتقر إلى وجود مفاهيم عربية الأصل، واستفيد من الأنطولوجيا في تلخيص آلي مقاد بالاستعلام، وذلك لتوسيع الاستعلام باستخدام العلاقات الدلالية في الأنطولوجيا، ولم يستفد من الأنطولوجيا في تلخيص عام غير مقاد بالاستعلام للاستفادة منها في تمثيل النصوص بمفاهيمها وعلاقاتها الدلالية، والاعتماد على هذا التمثيل في عملية التلخيص من خلال تثقيب الجمل وتحديد أهميتها اعتماداً على خصائص مفاهيمها وعلاقاتها الدلالية، مما قد يحسن من جودة التلخيص.

القسم العملي

تم إيجاد أنطولوجيا عربية اعتماداً على توسيع شبكة الكلمات العربية (Arabic WordNet (AWN وإغنائها بالاستعانة بخصائص الموسوعة العربية الغنية بالمعلومات والمفاهيم العربية في المجالات كافة وذلك باستخدام تقنية التقابل.

اقترح نظامان لتلخيص النصوص العربية هما: نظام تلخيص دلالي باعتماد الأنطولوجيا العربية المولدة بالتقابل، ونظام تلخيص إحصائي من أجل المقارنة والتقييم. وتم حساب درجة أهمية الجمل من خلال المجموع الخطي لخصائصها التي تضمنت: العبارات المفتاحية، التشابه مع العنوان، موقع الجملة، طول الجملة، العبارات البارزة، العبارات غير هامة. واعتمد في حساب هذه الخصائص في التلخيص الإحصائي اعتماداً على توارد المصطلحات فيها، بينما حسبت هذه الخصائص في التلخيص الدلالي اعتماداً على توارد مفاهيم الجمل التي تستنتج استناداً إلى العلاقات الدلالية في الأنطولوجيا العربية.