
بسم الله الرحمن الرحيم

تحليل النص باستخدام برنامج آر لطلاب الأدب

تأليف

Matthew L. Jockers

ترجمة

د. عبدالله عثمان الشعلان

الأستاذ المساعد بقسم علوم الحاسب

كلية علوم الحاسب والمعلومات

جامعة الملك سعود

د. هارون ناصر آل صقر

الأستاذ المشارك بقسم اللغة الإنجليزية

كلية العلوم والدراسات الإنسانية

جامعة الأمير سطام بن عبدالعزيز

دار جامعة الملك سعود للنشر

KING SAUD UNIVERSITY PRESS

ص.ب ٦٨٩٥٣ - الرياض ١١٥٣٧ المملكة العربية السعودية

ح دار جامعة الملك سعود للنشر، ١٤٤٤هـ (٢٠٢٢م)

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

جوكرز، ماثيو.

تحليل النص باستخدام برنامج آر لطلاب الأدب / ماثيو جوكرز؛ عبدالله عثمان الشعلان؛ هارون ناصر آل صقر - الرياض، ١٤٤٤هـ.

٢٥٥ ص؛ ١٧سم × ٢٤ سم

ردمك: ٧ - ٠٧١ - ٥١٠ - ٦٠٣ - ٩٧٨

١- البحث العلمي - معالجة البيانات أ. الشعلان، عبدالله عثمان (مترجم)

ب. آل صقر، هارون ناصر (مترجم) ج. العنوان

ديوي ٠٠١,٤٢٠٢٨٥ ١٤٤٤/٢٩٩

رقم الإيداع: ١٤٤٤/٢٩٩

ردمك: ٧ - ٠٧١ - ٥١٠ - ٦٠٣ - ٩٧٨

هذه ترجمة عربية محكمة صادرة عن مركز الترجمة بالجامعة لكتاب:

Text Analysis with R for Students of Literature
By: Matthew L. Jockers

وقد وافق المجلس العلمي على نشرها في اجتماعه الرابع للعام الدراسي ١٤٤٣هـ، المعقود بتاريخ ٥/ ٣/ ١٤٤٣هـ، الموافق ١٢/ ١٠/ ٢٠٢١م.

دار جامعة الملك سعود للنشر
KING SAUD UNIVERSITY PRESS

مقدمة المترجمين

الحمد لله رب العالمين والصلاة والسلام على حبيبنا وقدوتنا محمد بن عبدالله وعلى آله وصحبه ومن والاه إلى يوم الدين، أما بعد:

فإن هذا الكتاب " تحليل النص باستخدام برنامج آر لطلاب الأدب" يعرض كما هو ظاهر من عنوانه جوانب استخدام البرمجيات والتطوير التقني بعلوم الحاسب الآلي وتعلم الآلة في العلوم والأبحاث الإنسانية ومنها النصوص والكتب الأدبية، حيث يندرج هذا الكتاب ضمن مجال علوم حوسبة اللغة أو الإنسانيات الرقمية. تم تقسيم الكتاب إلى ثلاثة أبواب:

أولاً: يتناول الكتاب في الباب الأول مقدمة تعريفية حيال برنامج آر وكيفية استخدامه، وشرح العديد من الدوال التي تستخدم لمعالجة النصوص والتحليل على مستوى المفردات.

ثانياً: يقوم الكتاب بتدريب وتعليم الطلبة أو الباحثين في علوم اللغات والأدب بشكل خاص على كيفية تحليل النصوص على مستوى الجملة والمقاطع والنص الواحد، ويقدم شرحًا لأدوات متعددة في برنامج آر يستطيع من خلالها الباحث قياس تنوع المفردات وثراءها واستخراج الكلمات الأساسية الواردة في النص.

ثالثاً وأخيراً: يقوم الكتاب بالعمل على تدريب الباحثين أو طلاب وطالبات الأدب أو متخصصي اللغويات على كيفية التحليل على مستوى مجموعة من النصوص باستخدام وسائل تعلم الآلة كالتجميع والتصنيف ونمذجة الموضوعات وتصورها.

نستهدف بتقديم هذا الكتاب قراء العربية ومتخصصي علوم اللغة والأدب والمهتمين ببرمجيات اللغات الحية لمساعدتهم في دراستها وإجراء أبحاثهم المتعلقة بها، بحيث يتمكنوا من الاستفادة من هذه البرامج كبرنامج آر وغيرها. وحيث إن الأسلوب التقليدي في تحليل النصوص قد

يكون عائقاً أمام المتخصصين في الوصول لنتائج مرضية يمكن تعميمها، إلا أن هذه البرمجيات تساعد في فحص ملايين من الكتب والنصوص بشكل أفضل وأسرع، وهذا يساعد بشكل كبير في تجويد البحوث في العلوم الإنسانية. كما يُفيد هذا الكتاب متخصصي علوم الحاسب الآلي وبشكل خاص المهتمين بمجال تعلم الآلة والذكاء الصناعي للوصول للإنسانيات الرقمية Digital Humanities.

الكتاب مع صغر حجمه يتميز بلغته البسيطة والسهلة لمن لا يتقن البرمجة وخلفيته التقنية ليست كخلفية متخصصي البرمجة، إذ يأخذ الكتاب قارئه خلال أبوابه خطوةً بخطوة حتى يكتسب مهارات استخدام برنامج آر وتطبيقه على النصوص الأدبية واللغوية.

في الختام، نتمنى للقارئ تجربة ثرية وممتعة. ونرجو أن يكون في هذا الكتاب خدمة للغتنا العربية وإثراء للمحتوى العربي. والحمد لله رب العالمين.

المترجمين

إهداء المؤلف

إلى أمي،
التي تُفضّل اتباع التعليمات

شكر وتقدير

لعدة سنوات قدمت دورات في تحليل النص باستخدام مزيج من الأدوات ولغات البرمجة المختلفة. فدربت الطلاب على استخدام بيرل وبايثون وبي أتش بي وجافا وحتى إكس إل إس تي. ولتحليل البيانات الناتجة، استخدمنا غالبًا إكسل. وفي عام ٢٠٠٥ تقريبًا، بتحريك من كلوديا إنجل ودانييلا فيتن اللتين عرضتا عليّ بعض النصائح المفيدة على مستوى المبتدئين، بدأت في استخدام آر بدلاً من إكسل. ولفترة طويلة بعد ذلك لازلت أكتب الكثير من الشفرات لتحليل النص في بيرل أو بايثون أو بي أتش بي ثم استيراد النتائج إلى آر للتحليل. وفي عام ٢٠٠٨، قررت أن سير العمل هذا غير مستدام. فكنت أقضي الكثير من الوقت في نقل البيانات من بيئة إلى أخرى. فقررت أن أذهب إلى تركيا الباردة وأتخلى عن كل شيء لصالح أن أتعلم برنامج آر. ومنذ أن انتقلت، نادراً ما كان عليّ البحث في مكان آخر.

ولحسن الحظ، كما كنت أنتقل إلى استخدام آر، كان ينتقل إليه أيضًا الآلاف من الأشخاص الآخرين؛ فكان مجتمع الإنترنت من المبرمجين والمطورين للبرنامج يتوسع في اللحظة التي احتجت إليهم فيها بالضبط. وتعتبر موارد المساعدة على الإنترنت اليوم رائعة، ولولا ها ما تمكنت من كتابة هذا الكتاب. وهناك عدد هائل من المبرمجين وممن صنعوا بعض الحزم المفيدة بشكل لا يصدق. وذكرت حفنة صغيرة فقط من هذه الحزم في هذا الكتاب (وهذا، في نهاية المطاف، دليل للمبتدئين) ولولا أناس مثل ستيفان تي أتش. غريس، وهارالد باين، وهادلي ويكهام، لافتقر هذا الكتاب ومجتمع آر إلى الكثير. وأنا مندهش من مدى صداقة مجتمع آر في الإنترنت الذي أصبح مفيدًا وتسوده علاقات الود؛ لم يكن الأمر كذلك في السنوات الأولى، ولذلك أود أن أشكر الذين كتبوا الحزم وساهموا في مشروع آر وأيضًا جميع من قدموا المشورة بشأن منتديات آر، وخاصة في قائمة مساعدة-آر `R-help` وعلى موقع `stackoverflow.com`.

بدأ هذا الكتاب كسلسلة من التمارين للطلاب الذين كنت أعلمهم في جامعة ستانفورد؛ رأوا الكثير من هذا المحتوى في شكل خام وأقل صقلاً. وثم الكثير منهم ويجب أن نشكر كل فرد على حدة، لذا شكراً، جميع تلاميذي السابقين والحاليين، على سعة صدركم وتعليقاتكم. فلولاكم لأصبح هذا الكتاب، بغض النظر عن عيوبه وأوجه القصور فيه، عديم الفائدة تماماً.

قمت أولاً بتجميع المواد من فصولي إلى مخطوطة في عام ٢٠١١، ومنذ ذلك الحين قمت بمشاركة أجزاء من هذا النص مع بعض الزملاء. وأجرى ستيفان سنكلير اختبار من النسخة التجريبية لهذا الكتاب في دورة قام بتدريسها في جامعة مكغيل. وقدم هو وطلابه تعليقات قيمة. فقد قرأ مكسيم رومانوف معظم هذه المخطوطة في أوائل عام ٢٠١٣. ولقد قدم التشجيع والتعليقات وأقنعني في النهاية بتحويل المخطوطة إلى لاتكس LaTex من أجل تنضيد الحروف بشكل أفضل، وقادني هذا في النهاية إلى سويف Sweave ونيترknitr : رزمتان لآر سمحتا لي بتضمين وتشغيل نصوص آر البرمجية من داخل هذه المخطوطة ذاتها. ولذا، شكراً، مرة أخرى لمكسيم، وكذلك لمؤلفي سويف ونيتر وفيدريش ليستش وييهوي شيه.[1] وأود أيضًا أن أشكر أولئك الذين قاموا بتنزيل مُسوَّدة هذه المخطوطة التي نشرتها على موقعي في أغسطس ٢٠١٣ وأقدم تقديري لهم وكذلك لأولئك الذين كتبوا إليّ مع تعليقاتهم وتصحيحاتهم وهم مذكورون في صفحة المساهمين التالية.[2] وأخيرًا، أشكر ابني

(١) فيدريش ليستش.سويف توليد ديناميكي للتقارير الإحصائية باستخدام تحليل بيانات القراءة والكتابة. وولفان هاردلي وبيرند رونز، محرران، إحصائيات ٢٠٠٢- وقائع في الإحصاء الحسابي، الصفحات ٥٧٥-٥٨٠. فيزيكا فيرلاج، هايدلبرغ، ٢٠٠٢. المعيار الدولي للكتاب رقم ٩-١٥١٧-٧٩٠٨-٣. ييهوي شيه. المستندات الديناميكية مع آر ونتر. تشابمان وهول / اتفاقية حقوق الطفل. المعيار الدولي للكتاب رقم ٤٨٢٢٠٣٥٣٠-٩٧٨، ٢٠١٣.

(٢) تضمنت مسودة ما قبل النشر لهذا النص المنشورة عبر الإنترنت المراجعة المفتوحة للفقرتين التاليتين:

- أخيرًا، أود أن أشكركم على الحصول على نسخة من هذا النص الإلكتروني المطبوع مسبقاً وتصفحه. آمل أن تقدم لي تعليقاتك وأن تساعدني في جعل النسخة النهائية المطبوعة جيدة قدر الإمكان. وعندما تقدم هذه التعليقات، سأضيف اسمك إلى قائمة المساهمين لتضمينها في الإصدارات المطبوعة والإلكترونية. وإذا قدمت مساهمة كبيرة، فسأعترف بذلك على وجه التحديد.

- لم أتعلم آر بنفسي، ولا يزال هناك الكثير حول البرنامج يجب أن أتعلمه. وأريد أن أعترف بهاتين الواقعتين بصورة مباشرة وخاصةً الإشارة إليكم جميعاً، يا من اقتطعتم من وقتكم للمساهمة في هذه المخطوطة وجعل عالم آر مكانًا أفضل.

البالغ من العمر ١٤ عامًا الذي دخل ونفذ كل سطر من التعليمات البرمجية في هذا الكتاب. وعلى الرغم من أن هذا النص قد لا يكون سهلاً للغاية بحيث يمكن لصاحب الكهف استخدامه، فإنني متيقن أنه بإمكان طالب متوسط المستوى استخدامه دون الحاجة إلى تكثيف التدريبات.

المساهمون

نزلت مسودة هذا الكتاب أكثر من ١٠٠٠ مرة بعد نشره على موقعي في أغسطس ٢٠١٣. وقد قدم القراء المدرج أسمائهم أدناه تعليقات قيمة حول المخطوطة، وأشكرهم جميعاً على مساهماتهم في المخطوطة النهائية. وجميع مساهماتهم المحددة محفوظة على الموقع التالي:

http://www.matthewjockers.net/text-analysis-with-r-for-students-of-literature/

وكان أكبر عدد من المساهمات من تشارلز شيرلي، الذي قدم ١٣٣ تعليقًا. وجاءت أهم مساهمة في النصوص البرمجية من كارمن ماكو، الذي اكتشف خللاً بسيطاً جدًا في الفصل الرابع. شكراً لكم جميعًا.

١- بروتنوف، ميكال
٢- بينتيكوست، ستيفن
٣- تيدرو، كيمبرلي
٤- جونسون، بول
٥- شيرلي، تشارلز
٦- شيه، ييهوي
٧- فرانكوم، جيريد
٨- كوماري، أشانكا
٩- لودون، جون
١٠- ماكمولين، كيفن
١١- ماكو، كارمن

١٢- ماينر، ماثيو ج.
١٣- هوبر، ألكسندر
١٤- هوك، براندون
١٥- وولف، مارك
١٦- ويروين، أوستن

تمهيد

هذا الكتاب مقدمة في تحليل النص الحسابي باستخدام لغة البرمجة مفتوحة المصدر آر. وعلى عكس الكتب الأخرى الجيدة التي تتناول استخدام آر في التحليل الإحصائي للبيانات اللغوية[1] أو في إجراء لغويات المتون المُجمَّعة الكمية،[2] يختص هذا الكتاب بطلاب الأدب والباحثين فيه، ثم يستهدف بصفة عامة المتخصصين في العلوم الإنسانية الذين يرغبون في توسيع مجموعة أدواتهم المنهجية لتشمل المنهج الكمي والحسابي في دراسة النص. وأردت أن يكون هذا الكتاب أيضا وجيزاً وفي صلب الموضوع. آر برنامج معقد لا يمكن لأي كتاب مدرسي تبسيطه. وينصب التركيز هنا على جعل التقنية مستساغة والأهم من ذلك جعلها مفيدة ومجزية على الفور! وأقصد بالمكافأة هنا ليس الشعور بالرضا الذي يجنيه المرء من إتقان لغة البرمجة، بل تحديداً العائد السريع من استثمارك. فستبدأ في تحليل النص ومعالجته على الفور، وسيرشدك كل فصل إلى تقنية أو عملية جديدة.

يوفر الحساب إمكانية الوصول إلى المعلومات في النصوص التي لا يمكننا ببساطة جمعها باستخدام طرقنا النوعية التقليدية في القراءة الدقيقة والتوليف البشري. وتأتي المكافأة في التمكن من الوصول إلى تلك المعلومات على الصعيدين الكلي والجزئي. وإذا نجح هذا الكتاب، فيمكنك الانتهاء منه بأساس، مع عرض واسع للتقنيات الأساسية وفهم أساسي للاحتمالات. وسيبدأ التعلم الحقيقي عند وضع هذا الكتاب جانباً وإنشاء مشروعك الخاص. فهدفي هو منحك خلفية كافية حتى تتمكن من البدء في هذا المشروع براحة تامة ولكي تكون قادراً على مواصلة التعلم وتثقيف نفسك.

(١) باين، هـ.أ. تحليل البيانات اللغوية: مقدمة عملية في الإحصاء باستخدام آر، مطبعة جامعة كامبردج، ٢٠٠٨.

(٢) جريس، ستيفان ث. علم اللغة الكمي مع آر: مقدمة عملية. نيويورك: روتليدج، ٢٠٠٩.

عندما أناقش عملي كإنساني مختص بمجال الحوسبة، أُسأل مرارًا وتكرارًا عما إذا كانت الطرق والمناهج التي أؤيدها تنجح في جلب معرفة جديدة لدراستنا للأدب. وجوابي هو "نعم" بصوت جازم وعال. في الوقت نفسه، يجب أن تكون كلمة "نعم" مستحقة بعض الشيء؛ ليس كل ما يكشفه تحليل النص هو اكتشاف مذهل. فيهدف قدر كبير من العمل الحسابي على وجه التحديد إلى اختبار المعرفة التي نعتقد أننا نمتلكها بالفعل أو رفضها أو إعادة تأكيدها. وخلال محاضرة عن الأنماط الكلية للأسلوب الأدبي في رواية القرن التاسع عشر، استخدمت مثالاً من رواية موبي ديك. ولقد أظهرت كيف أن رواية موبي ديك متغيرة إحصائياً في مجموعة تضم ١٠٠٠ رواية أمريكية أخرى من القرن التاسع عشر. فرفع أحد الزملاء يده وأشار إلى أن علماء الأدب يعرفون بالفعل أن رواية موبي ديك رواية مضلة، وتساءل، لما العناء في حساب إجابة عن سؤال نعرفه بالفعل؟

كان سؤال زميلي وجيهاً كاشفًا أيضاً. فكشف السؤال الكثير عن تقاليدنا العلمية في العلوم الإنسانية. إنه، في الوقت نفسه، سؤال مثير للسخرية: كتخصص، كنا نميل إلى تأييد فكرة أن الحجج الأدبية لا تنتهي أبدًا. فهل نعرف حقاً أن رواية موبي ديك أضلولة؟ فربما تكون الرواية مجرد قصة غريبة مقارنةً بالروايات الأمريكية العشرين أو الثلاثين الأخرى التي تعودنا على دراستها إلى جانبها. ولم تكن وجهة نظري في استخدام الرواية هو التظاهر بأنني اكتشفت شيئًا جديدًا حول موقف الرواية في التقاليد الأدبية الأمريكية، بل لتقديم نوع جديد من الأدلة ومنظور جديد لهذه المسألة وبهذا أعزز (في هذه الحالة) الفرضية الحالية.

وإن حدث أن أكّد نوع جديد من الأدلة على ما توصلنا إليه من اعتقاد في استخدام مناهج أكثر حدسية بشكل كبير، ألا ينبغي النظر إلى هذا الدليل الجديد على أنه شيء جيد؟ إذا عادت أحدث مركبة روفر من المريخ بمزيد من الأدلة على أن الكوكب كان يمكن أن يدعم الحياة ذات مرة، فسيكون ذلك دليلًا جديدًا مهمًا. إلا أن الأمر لن يكون مذهلاً أو مثيرًا للاهتمام كأول اكتشاف للميكروبات، أو أول اكتشاف للجليد على سطح المريخ، إلا أنه سيكون دليلًا مهمًا، لكنه سيضيف قطعة أخرى إلى لغز أكبر. فلذلك، يمكن للنُهج الحسابية للدراسة الأدبية أن تقدم أدلة تكميلية، وأعتقد أن هذا شيء جيد.

وتنطوي الأساليب الموضحة في هذا الكتاب أيضًا على إمكانية تقديم أدلة متناقضة أو أدلة تتحدى نظرياتنا التقليدية أو الانطباعية أو القصصية. وبهذا المعنى، توفر لنا الطرق بعض الفرص لنوع

التزوير الذي يقدمه كارل بوبر وما بعد الفلسفة الوضعية عامةً كحل وسط بين الوضعية الصارمة والنسبية الصارمة. ولكن نظرًا لأن هذه الأساليب يمكن أن تخلق تناقضًا، يجب ألا ننشغل في لعبة الأرقام حيث نقدر فقط الأفكار القابلة للاختبار. فبعض التفسيرات تصلح للاختبار الحسابي أو الكمي؛ والبعض الآخر لا يصلح، وأعتقد أن هذا شيء جيد.

وأخيرًا، يمكن أن تؤدي هذه الطرق إلى اكتشافات جديدة حقًا. فتحليل النص الحسابي لديه طريقة لإبراز تفاصيل وخصائص النصوص التي نفتقدها بالعين المجردة.[٣] وباستخدام التقنيات الحسابية فقط، اكتشف باتريك جولا مؤخراً أن ج. ك. رولينج هي المؤلفة الحقيقية لرواية "نداء الوقواق"، فلقد كتبت الكتاب تحت الاسم المستعار روبرت غالبريث. وبطبيعة الحال، أعتقد أن اكتشاف جولا أمر جيد للغاية.

وهذا كل ما أود قوله بخصوص نظرية أو علة تحليل النص. أما في كتابي الآخر، فأنا أكثر جدلاً بعض الشيء.[٤] المهمة هنا ليست الدفاع عن الأساليب، بل مشاركتها.

لينكولن، نيو إنجلاند — ماثيول. جوكرز

يناير ٢٠١٤

(٣) انظر فلاندرز، جوليا. "التفصيل والنصوص الرقمية ومشكلة التحجيم." تقنية النصوص، ٢: ٢٠٠٥، ٤١–٧٠.

(٤) جوكرز، ماثيو. التحليل الكلي: الأساليب الرقمية والتاريخ الأدبي. مطبعة جامعة إلينوي، ٢٠١٣.

المحتويات
CONTENTS

الباب الثالث: التحليل الكلي