قياسات المعلومات*

جان تاجيو ستكليف
ترجمة/ عبد الرحمن فراج
مدرس مساعد – قسم المكتبات والوثائق
جامعة القاهرة– فرع بني سويف
aafarrag@hotmail.com

لم يعد مصطلح قياسات المعلومات Informetrics شائعاً بين علماء المعلومات فى أوربا الغربية وأمريكا الشمالية إلا فى الأعوام الخمسة الماضية فحسب، وذلك كمجال عام ينضوى تحته كل من المجالين السابقين: القياسات الوراقية Bibliometrics وقياسات النشاط العلمى Scientometrics . ويعود تاريخ قبول هذا المصطلح ـ فى الحقيقة ـ إلى المؤتمر الدولى عن القياسات الوراقية والجوانب النظرية لاسترجاع المعلومات المنعقد فى دبنبيك ببلجيكا، حيث اقترح بى سى بروكس B.C. Brookes أن يشتمل عنوان المؤتمر التالى، المحدد له أن ينعقد فى لندن بكندا عام ١٩٨٩، على مصطلح قياسات المعلومات. على ذلك، جاء الاجتماع التالى معنوناً بالمؤتمر الدولى للقياسات الوراقية وقياسات النشاط العلمى وقياسات المعلومات. أما عنوان الاجتماع الثالث فى هذه السلسلة، والذى انعقد فى بنجالور بالهند عام ١٩٩١، فقد أشار إلى القبول النهائى لهذا المصطلح: «المؤتمر الدولى لقياسات المعلومات».

ونقلاً عن بروكس، فإن مصطلح قياسات المعلومات أقترح للمرة الأولى فى ألمانيا الغربية عام ١٩٧٩، وذلك على يد أوتوناك (Brookes, 1990). فيما قدمت لجنة FID أهدافاً محددة أكثر اتساعاً للمصطلح، فى تعريفها له فيما بعد، وذلك بربطه بالبيانات البحثية والتقنية. وعلى أية حال، فإن استخدام المصطلح، بالمعنى الذى يُستخدم به الآن، لم يكن متفقاً عليه بصورة كافية.

إن القياسات الوراقية Bibliometrics هى دراسة الجوانب الكمية لإنتاجية المعلومات المسجلة، وبثها، والإفادة منها. ولقد تم ابتكار وتطوير عديد من النماذج والمقاييس الرياضية لأجل هذه العمليات، وتم استخدام هذه النماذج والمقاييس ـ من ثم ـ فى التنبؤ واتخاذ القرار. ولقد ظهر هذا المصطلح للمرة الأولى فى مقال لبريتشارد (١٩٦٩)، وذلك كبديل عن المصطلح السابق «الوراقة الإحصائية». ويُعرِّف بعض الباحثين، خاصةً وايت ومكاين (١٩٨٩) القياسات الوراقية بأنها دراسة مفردات

* Tague - Sutcliffe, Jean. An Introduction to Informetrics. Information Processing & Management. vol. 28, no. 1 (1992). pp. 1 - 3.

* والمقال فى الأصل مقدمة لهذا العدد الخاص من دورية Inf. Proc. & Manag. عن قياسات المعلومات.

الإنتاج الفكرى كما تنعكس فى الوراقيات. فيما يرى بروكس (١٩٩٠) أنها الآن أوثق صلةً بدراسات المكتبات.

وكان أوضح ظهور لمصطلح قياسات النشاط العلمى Scientometrics على عنوان دورية أُنشئت على يد تى بروان T. Braun عام ١٩٧٧، حيث صدرت أصلاً فى المجر، وتُنشر الآن فى أمستردام. وقياسات النشاط العلمى هى دراسة الجوانب الكمية للنشاط العلمى باعتباره مجالاً أو نشاطًا اقتصاديًا. وتُعد بذلك جزءاً من علم اجتماع المعرفة، كما أن لتطبيقاتها دوراً فى رسم السياسة العلمية. وتحيط هذه القياسات بالدراسات الكمية للأنشطة العلمية، والتى تشمل ـ ضمن ما تشمل ـ على الإنتاج الفكرى المطبوع. وعلى ذلك، فإنها تتداخل وتتشابك، إلى حدٍ ما، مع القياسات الوراقية.

أما قياسات المعلومات فهى دراسة الجوانب الكمية للمعلومات فى أى شكل لها ـ وليس فحسب فى الوراقيات أو التسجيلات الوراقية، وفى أى قطاعٍ اجتماعى ـ وليس فقط بين المشتغلين بالبحث العلمى. وعلى ذلك، فهى تعنى بالجوانب الكمية للاتصال غير الرسمى أو الشفهى جنباً إلى جنب الاتصال الوثائقى، كما تعنى بالجوانب الكمية للحاجة إلى المعلومات والإفادة منها، ليس المنتقى منها فحسب بل والمعلومات الثانوية أيضاً. أى أن بإمكان هذه القياسات احتواء والإفادة من وتوسيع عديد من دراسات مقاييس المعلومات التى تقع خارج حدود كل من القياسات الوراقية وقياسات النشاط العلمى.

إن لقياسات المعلومات إطاراً عملياً وآخر نظريا. والاهتمامات الأساسية فى هذا المجال تؤكد على تطوير النماذج الرياضية، فيما تنصب الاهتمامات الثانوية على اشتقاق مقاييس للظواهر العديدة محل الدرس. وقيمة أى نموذج من هذه النماذج تكمن فى قدرته على التوسل بأقل قدر من المعاملات لحصر خصائص مجموعات شتى من البيانات، مثل: الصغية العامة، والتمركز، والتشتت، والطريقة التى تتغير بها البيانات بمرور الزمن. فضلاً عن ذلك، فإن هذه النماذج تتيح لنا التنبؤ بالسلوكيات فى المستقبل، وعزل الآثار التى يمكن أن تُحدثها العوامل المختلفة على المتغيرات محل الاهتمام. وهكذا، فإن هذه النماذج ـ متحدةً مع المقاييس التى تُشتق منها ـ تقدم لنا أساساً راسخاً لاتخاذ القرار فى الواقع العملى.

وبالرغم من أن مجالات تطبيق قياسات المعلومات فى الواقع العملى متسعة للغاية، إلا أن علماء القياسات الوراقية وقياسات النشاط العلمى قد ركزوا دراساتهم ـ فيما مضى ـ على النماذج والمقاييس الرياضية فى مجالات محدودة ومحددة جيداً، وهى المجالات التالية:

* الجوانب الإحصائية لتردد كل من اللغات، والكلمات، والعبارات؛ وذلك فى كل من نصوص اللغة الطبيعية والكشافات، وكذلك فى كل من الوثائق المطبوعة والوسائط الإلكترونية.

* خصائص المؤلفين ـ ونعنى بها إنتاجيتهم التى يمكن قياسها عن طريق عدد المقالات، أو غيرها من الوسائط؛ ثم مدى التعاون فيما بين هؤلاء المؤلفين.

* خصائص المصادر الوثائقية؛ وأكثر المصادر

خضوعاً للدراسة فى هذا الصدد توزيع المقالات بين مجموعات الدوريات فى تخصص موضوعى معين.

* تحليل الاستشهادات المرجعية، وتوزيعها بالنسبة للمؤلفين، والمقالات، والمؤسسات، والدوريات، وأماكن النشر؛ واستخدامها فى عمليات التقييم؛ ثم المصاحبة الوراقية التى تقوم على رسم خريطة شاملة للتخصصات العلمية.

* مدى الإفادة من المعلومات المسجلة أو الوثائقية: الإعارة الخارجية والداخلية للكتب فى المكتبات، والإفادة من الدوريات، والإفادة من مراصد البيانات.

* التعطل فى الإنتاج الفكرى، كما يتم قياسه من خلال مدى الإفادة من المكتبات أو من خلال تحليل الاستشهادات المرجعية.

* معدلات نمو الإنتاج الفكرى المتخصص، ومراصد البيانات، والمكتبات، ذلك النمو المصاحب ـ بالطبيعة ـ لتطور المفاهيم الجديدة فى التخصص الموضوعى.

وثمة ظاهرتان لم تتضحا، فى الماضى، كجزء من القياسات الوراقية أو قياسات النشاط العلمى، ولكنهما تتوافقان تماماً مع قياسات المعلومات، وهما:

* تحديد وقياس المعلومات.

* أنماط وخصائص مقاييس الأداء فى عمليات الاسترجاع.

والظاهرة الأولى تتم دراستها بالفعل فى مجالى الهندسة الكهربائية والرياضيات منذ ظهور نظرية شانون وويفر (١٩٤٩). والحق أن هناك صور معدلة من المقياس الأصلى لشانون وويفر قد تم تطويرها فى مجالات عديدة. وثمة عمل حديث يقدم فيه لوزى (١٩٩٠) عرضاً ممتازاً لهذا الجانب من قياسات المعلومات. أما مقاييس الأداء فى عمليات الاسترجاع فقد تمت دراستها على أيدى منظرى استرجاع المعلومات فى مجالى علم المعلومات وعلم الحاسبات. وقد تم عرض كل من الجوانب النظرية والتطبيقية لهذا الموضوع فى سلسلة المؤتمر السنوى الموسوم «المؤتمر الدولى للبحوث والتطوير فى مجال استرجاع المعلومات».

ويمكننا تحديد ثلاثة أسماء ارتبطت بالتطور المبكر للمجال الذى نطلق عليه اليوم قياسات المعلومات، وهم : لوتكا، وزيف، وبرادفورد. ولقد كان جميع هؤلاء الباحثين مهتمين تقريباً بظاهرة واحدة*. فقد اختص لوتكا بإنتاجية المؤلفين، واتجه زيف إلى دراسة تردد الكلمات، فيما ركّز برادفورد على إنتاجية الدوريات. وبرغم ذلك، وإضافةً إلى ما سبق، فقد مهّد هؤلاء الرجال للأساليب النظرية الأساسية فى المجال.

فقد قدم لوتكا (١٩٢٦) الأسلوب الأول لتوزيع التردد الطبقى للمفردات (والتى يمثلها فى هذه الحالة بحوث الكيمياء) عبر مجموعة من المصادر (التى يمثلها هنا المؤلفون)؛ ذلك الأسلوب الذى أصبح معروفاً فيما بعد بقانون التربيع العكسى.

أما زيف (١٩٤٩) فمن خلال دراسته لتردد

* يمكن القول ـ بصورة أدق ـ إن التوزيعات الثلاثة لكل من لوتكا وزيف وبرادفورد ذات أساس واحد تقريباً، إن كان كل منها يعتمد على نوع مختلف من البيانات. أى أن الرجال الثلاثة ينظرون إلى ظاهرة واحدة، ولكن كلاً من جانب مختلف (المترجم).

الكلمات فى مجموعة مختلفة من المصادر، قام بتطوير كل من أسلوبى توزيع التردد الطبيعى وتوزيع التردد الطبقى، وذلك على طريق توزيع رموز الكلمات وفقاً لطبقاتها. وهذا التوزيع الأخير (الذى هو تردد أو تكرار الكلمات عكسياً بما يتناسب وطبقة كل كلمة) هو ما أصبح يُعرف بقانون زيف.

ولقد أسهم برادفورد بأسلوبين نظريين*، أُطلق على كل منهما، بصورة أدت إلى الالتباس إلى حدٍ ما، قانون برادفورد. لقد ابتكر برادفورد الأسلوب التراكمى اللوغاريتمى الخطى لتوزيع التردد الطبقى (وذلك لتوزيع المقالات فى تخصص موضوعى معين عبر مجموعة الدوريات ذات الصلة). إضافةً إلى ذلك، قدَّم برادفورد فكرة المتوالية الهندسية لتمثيل العدد المتزايد من الدوريات فى قطاع النواة والقطاعات التالية له فى المجال الموضوعى المتخصص؛ حيث تشتمل كل من النواة والقطاعات لأخرى على عددٍ متساوٍ من المقالات، فيما تتناقص المقالات باستمرار من دورية إلى أخرى.

ولقد قام الباحثون المحدثون، وأبرزهم بروكس (١٩٦٨)، وليمكولر (١٩٦٧)، ومندلبرت (١٩٦١)، بتعميم هذه النماذج وإنشاء العلاقات فيما بينها. ولقد اضطلع بروكس، بشكل خاص، بالتأكيد على أهمية الأساليب اللوغاريتمية والطبقية فى نمذجة المعلومات. فيما حاول برايس Price (1976) إرجاع هذه الأساليب وغيرها إلى الخصائص العامة لما يُسمى بظاهرة «النجاح يلد نجاحاً». كما قام هايتن (١٩٨٢) بتقديم ما يمكن تسميته بأنه الوصف الأكثر عمومية للأساليب الملائمة للظاهرة محل قياسات المعلومات، وذلك فى وصفه لتوزيع non - Gaussian.

وفى الوقت الحاضر، يتجه مجال قياسات المعلومات بقوة أكثر اطراداً ناحية الجوانب النظرية والتطبيقية، وذلك عما كان عليه الحال فى عهد ريادة لوتكا وزيف وبرادفورد. فلقد وضع المنظِّرون نماذج وأساليب قياسات المعلومات على أسس رياضية وإحصائية أكثر رسوخاً عما كانت عليه فى الماضى. ويقدم لنا هذا العدد أمثلة متعددة من هذا التطور. إذ يقدم سيشل Sichel عرضاً واضحاً لما يمكن استخدامه من أسلوب عام للغاية، هو توزيع GIGP، للتطبيق على التوزيعات العديدة لظاهرة قياسات المعلومات. كما يبحث ثلاثة باحثون – هم روسو، وإيجى، وبرل – فى الخواص الرياضية لمقياس التمركز ومؤشر Gini. أما جلانزل Glanzel فيفحص الخواص الامبريقية والرياضية لمؤشر سرعة الاستشهادات المرجعية. ويبحث تابا Tabah فى الديناميات غير الخطية ونظرية الشوشرة كأساليب لدراسة نمو الإنتاج الفكرى. كما يفحص بوكشتاين Bookstein مدى ملائمة الأسلوب اللوغاريتمى الخطى لدراسة البيانات النوعية فى مجال قياسات المعلومات. وفى كل الأحوال، كانت التطبيقات ممثلةً فى هذه المقالات، بيد أن التركيز كان على التطورات والإسهامات الرياضية.

وعلى الطرف الآخر، ثمة مقالات عديدة تصب

* الواقع أن براد فورد قدَّم أفكاره الأساسية فى صيغتين: صيغة نظرية وتُدعى أيضا الصيغة اللفظية، وصيغة تطبيقية أو عملية وهى الصيغة البيانية (المترجم).

اهتمامها على البحث التفصيلى لتطبيقاتٍ بعينها. فيبحث كل من بيربام وتى إيه بروكس وآر إيه بروكس مدى تزايد وتناقص المصطلحات الموضوعية فى الإنتاج الفكرى فى موضوع الإيدز. أما باو Pao فيعرض للعلاقة ما بين التمويل ودرجة التعاون، وذلك اعتماداً على كمٍ هائلٍ من البيانات فى موضوع البلهارسيا. كما يدرس بونزى Bonzi التغير فى إنتاجية الأكاديميين بجامعة واحدة بمرور الزمن. ويجمل وولفرام توزيعات قياسات المعلومات ذات الصلة باسترجاع المعلومات اعتماداً على مراصد البيانات المكشفة، ويبين بواسطة المحاكاة كيف يمكن استخدام هذه النماذج لتحديد البناء الأمثل للملف، باسطاً الخصائص المميزة لكل من ملفات مراصد البيانات وملفات الاستفسارات. وفى جميع الأحوال، كانت المقالات فى هذا القسم ذات منحىً نظرى، بيد أن الاهتمام كان منصباً على التطبيقات.

لقد قدّم هذا التمهيد استعراضاً خاطفاً لمجال قياسات المعلومات. والمقالات اللاحقة له تواً سوف توسّع من هذا الاستعراض من عدة وجوه. ولمزيد من المعلومات عن هذا المجال سريع التطور، فإننا ننصح بوقائع المؤتمرات الدولية آنفة الذكر، وبالفصلين اللذين وردا فى المراجعة العلمية السنوية لعلم المعلومات وتقنياتها *Annual Review of Information Science and Technology* عن القياسات الوراقية. وتقدم قائمة المراجع المختارة الملحقة بهذا المقال، بيانات وافية عن هذه الأعمال، فضلاً عن الإشارة إلى بعض البحوث الرائدة فى المجال.

مراجع مختارة

Bradford, S.C. (1934). Sources of information on specific subjects. *Engineering, 137*, 85-86.

Brookes, B.C. (1968). The derivation and application of the bradford-zipf distribution. *Journal of Documentation, 24*(4), 247-265.

Brookes, B.C. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about? In *Informetrics 89/90*. Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (Eds.) (1988). *Informetrics 87/88; select proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25-28 August 1987*. Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (Eds.) (1990). *Informetrics 89/90; selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics, London, Ontario, Canada, 5-7 July 1989*. Amsterdam: Elsevier.

Haitun, S.D. (1982). Stationary scientometric distributions. *Scientometrics, 4*(1), 5-25; *4*(2), 89-104; *4*(3), 181-194.

Leimkuhler, F.F. (1967). The Bradford distribution. *Journal of Documentation, 23*(3), 197-207.

Losee, R.M. (1990). *The Science of Information: Measurement and Application*. San Diego, Calif.: Academic Press.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science, 16*(12), 317-323.

Mandelbrot, B. (1961). On the theory of word frequencies and on related Markovian models of discourse. In *Structure of Language and its Mathematical Aspects: Proceedings of Symposia in Applied Mathematics* (p. 12).

Narin, F., & Moll, J. (1977). Bibliometrics. *Annual Review of Information Science and Technology, 12*, 35-58.

Price, D.J. de S. (1976). A general theory of bibliometric and other cumulative disadvantage processes. *Journal of the American Society for Information Science, 27*(5), 292-306.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*, 348-349.

White, H.D., & McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology, 24*, 119-186.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Reading, MA: Addison-Wesley.