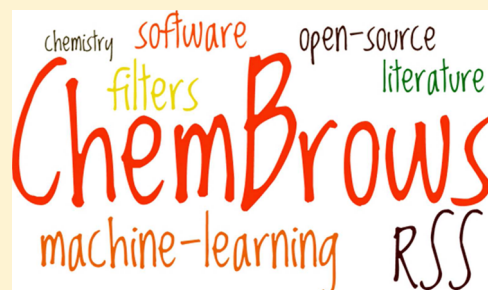


ChemBrows: An Open-Source Application Software To Keep Up to Date with the Current Literature

Jean-Patrick Francoia and Laurent Vial*

Institut des Biomolécules Max Mousseron, UMR 5247 CNRS—Université de Montpellier—ENSCM, Place Eugène Bataillon, 34296 Montpellier cedex 5, France

ABSTRACT: Working as an enhanced RSS reader that integrates keyword-based filters and a machine-learning-based recommendation engine, ChemBrows is a software that will significantly help scientists/teachers/students to tame the flood of publications. ChemBrows is available on multiple platforms as a free and open-source software at www.chembrows.com.



KEYWORDS: Graduate Education/Research, Computer-Based Learning, Chemoinformatics, Interdisciplinary/Multidisciplinary

INTRODUCTION

To stay up-to-date with the latest research results as a scientist/teacher/student has clearly become a burden. According to Thomson Reuters Web of Science, the number of publications increased linearly over the last two decades, with in 2014 a stunning figure of +600 papers published every single day in the sole field of chemistry (Figure 1, left).¹

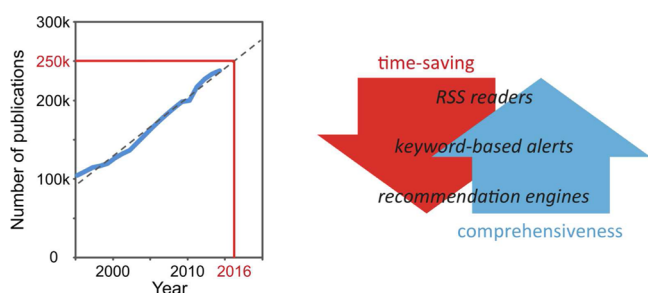


Figure 1. Left: Number of publications in the chemical sciences over the last two decades (see ref 1). Right: Available approaches toward the survey of the literature and their time-saving/comprehensiveness trends.

Following this growing rate, the milestone of a quarter million publications per year should be reached this year. Multiple factors may contribute to this rise in research output, which includes the growth of universities internationally, the emergence of new areas of research, and the infamous pressure to “publish or perish”. Though the two former factors are good news, the latter one certainly contributes to the avalanche of tedious—even fraudulent—papers in literature.² As a consequence, finding new, relevant, and inspiring research work is looking for a needle in a haystack.

COMMON METHODOLOGIES TOWARD THE SURVEY OF THE LITERATURE

Various methodologies can be applied to tackle this issue. Among them, the old-fashioned literature routine involves the regular browsing of publishers’ Web sites or RSS feeds to pick out fresh papers of interest.³ Although this methodology is nowadays time-consuming in regard to the flood of publications, it still has the advantage to be the most comprehensive one. In addition, this routine leads naturally to the browsing of papers from various research fields, which may stimulate unexpected connections and bracing ideas. Less fastidious solutions are available for literature surveying.

By using keywords-based alerts as provided for instance by PubMed, it is possible to significantly reduce the input of papers. Such an approach is almost mandatory in order to efficiently follow its own topic(s) of interest. On the other hand, a survey based on keywords potentially introduces blinkers to what is scientifically possible.

More recently, literature-recommendation services have emerged (e.g., ReadCube and Mendeley, to name but two), relying on algorithm-based engines that learn from their users’ interests to suggest relevant contents.⁴ These tools often provide with unnoticed and pertinent research articles, but—as a balance—also narrow the flood of literature to a trickle. To summarize, these previous methodologies toward the survey of literature come with a distinct time-saving/comprehensiveness ratio, with each methodology having its own advantages and being indeed complementary to the others (Figure 1, right).

Received: January 12, 2016

Revised: April 18, 2016

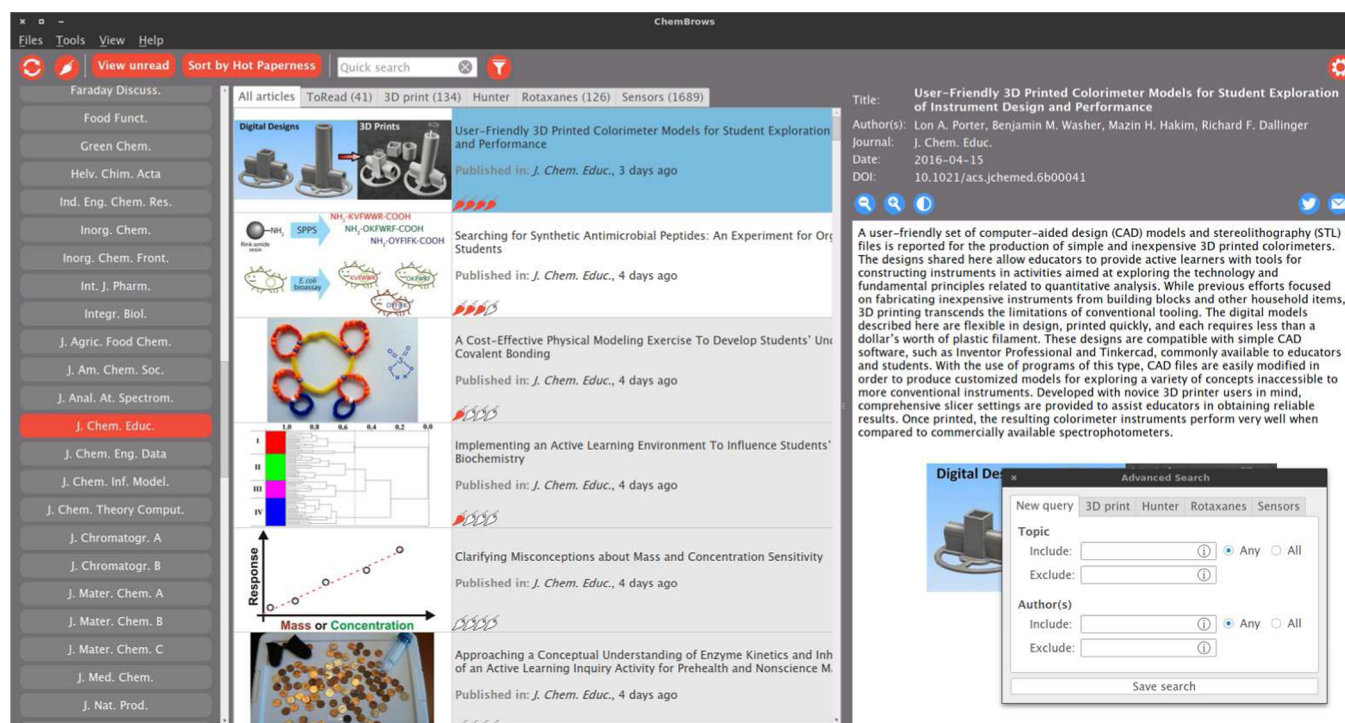


Figure 2. Interface of ChemBrows.

Herein, we introduce ChemBrows, an in-house piece of software that proposes an integrated solution to stay up-to-date with the literature.

■ AN ALL-IN-ONE SOLUTION WITH CHEMBROWS

Working primarily as a RSS reader, ChemBrows feeds the user with the latest papers from a customizable selection of journals (Figure 2).

A catalogue of 122 journals is included so far, but virtually any title can be further added, including from other disciplines.⁵ A dock appears upon mouse hovering on the left-side of the interface, allowing to display/hide titles (Figure 2, left side of the screenshot). Because the browsing of papers is significantly facilitated by the presence of graphics, ChemBrows proposes—unlike classical readers—the graphical abstract be embedded within the entries (Figure 2, middle of the screenshot). Once an entry is selected, the corresponding untruncated abstract is displayed (Figure 2, right side of the screenshot), and double-clicking on an entry opens the corresponding article's landing page in the user's default web browser. Users can also bookmark entries that will be saved in a separate "ToRead" tab.

In addition to providing scientists/teachers/students with the latest literature from a personal selection of journals, ChemBrows allows the creation of savable keyword-based filters in order to follow favorite topics and/or authors (Figure 2, dialogue box in the bottom-right corner of the screenshot). It is obviously possible to perform multiple keywords searches that involve Boolean operators.

Because it also includes a machine-learning algorithm, ChemBrows can classify the entries by "hot paperness" (Figure 2, hot peppers in the screenshot), which is a sorting based on a support-vector network (Figure 3).⁶

To put it simply, when the read entries are—or are not—liked, the algorithm extracts the words from the abstracts and titles and gives them a score. Further incoming abstracts are

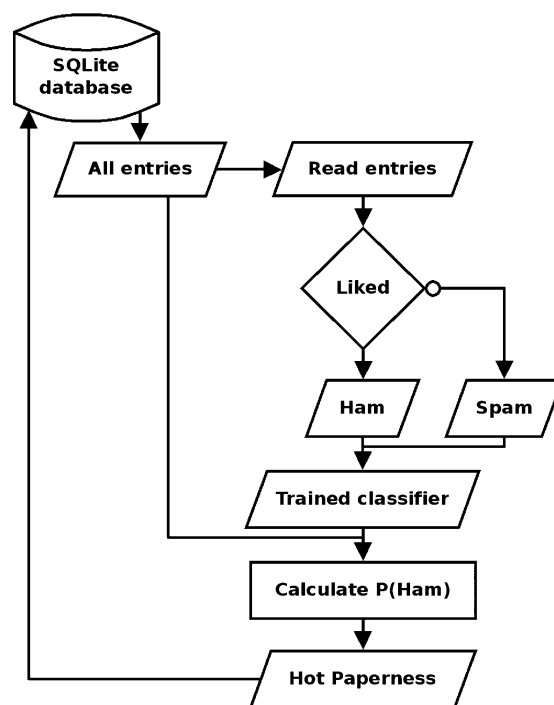


Figure 3. Simplified flowchart representing the classification by "hot paperness" with a support-vector network.

evaluated with respect to its words' score, and are subsequently labeled with one to four hot peppers as a prediction of its significance for the user. Obviously, the more the user feeds ChemBrows, the more ChemBrows feeds back the user with pertinent articles.

ChemBrows is running on multiple platforms (i.e., Linux, Mac OS X, and Windows)^{7,8} and is distributed under a GNU General Public License.⁹ The program is written in Python 3

and relies on numerous modules and packages.¹⁰ Information is parsed from the free content of the publishers' Web sites (i.e., RSS feeds, and possibly landing pages if graphical abstracts and/or full abstracts are missing) and then stored locally in a SQLite database, which can be exported or imported from one computer to another. Upon the first boot, the software requires an e-mail address, the only personal information collected, that is only used to deliver important news/update notifications. The graphical user interface was designed to be intuitive, a short embedded tutorial allowing the easy mastering of every ChemBrows' features.

CONCLUSION

An all-in-one enhanced RSS reader, keyword-based filter, and recommendation engine, ChemBrows offers to the user an unprecedented versatility toward the survey of the literature. We would like to emphasize that ChemBrows is an open-source software that is dedicated to the community of chemists and has noncommercial purpose. Because there are certainly multiple ways to improve upon this software, the source code of ChemBrows is fully available and contributions—including forks toward other disciplines—are enthusiastically welcome.

AUTHOR INFORMATION

Corresponding Author

*E-mail: laurent.vial@univ-montp2.fr.

Author Contributions

L.V. and J.-P.F. conceived the software. J.-P.F. carried out the programming work. L.V. wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are very grateful to Timothy C. Eichler and Ryan Kelly for their valuable contributions during the debugging phase of ChemBrows. We also thank Joseph Chamieh, Sébastien Clément, Michael Smietana (Université de Montpellier), and Piotr Nowak (University of Groningen) for beta-testing ChemBrows and for their helpful comments.

REFERENCES

- (1) Data were generated from <http://apps.webofknowledge.com>. Typical search input: databases = all, year published = 2014, refined by research areas = chemistry.
- (2) Lutz, J.-F. Slow science. *Nat. Chem.* **2012**, *4*, 588–589.
- (3) Pence, L. E.; Pence, H. E. Accessing and Managing Scientific Literature: Using RSS in the Classroom. *J. Chem. Educ.* **2008**, *85*, 1449–1452.
- (4) Gibney, E. How to tame the flood of literature. *Nature* **2014**, *513*, 129–130.
- (5) New titles will be added upon requests from users on ChemBrows' Web site at <http://www.ChemBrows.com/website/index.php?contact> (accessed Apr 2016); Twitter account at @ChemBrows; or GitHub repository at <https://www.github.com/chembrows/ChemBrows> (accessed Apr 2016).
- (6) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (7) The installation files of ChemBrows are available at <http://www.chembrows.com>.
- (8) After downloading the program, it may be installed on Windows by double-clicking on the executable file (setup ChemBrows 0.9.6 (32bit).exe). Mac OS X and Linux users should unzip the archive files

(tar.bz2), and then execute the package file (ChemBrows.pkg) or launch the executable file (gui), respectively.

(9) As specified by the Free Software Foundation, see: <http://www.gnu.org/licenses/gpl-3.0.en.html>.

(10) The source code of ChemBrows is available at <https://www.github.com/chembrows/ChemBrows> (accessed Apr 2016).