

Best practises for chemical structure registration and searching in relational databases

Christoph Steinbeck^{*1}, Wolf-Dietrich Ihlenfeldt², Stefan Kuhn¹, Paula de Matos¹, Jerome Pansanel³, Mark Rijnbeek¹, Ernst-Georg Schmid.⁴

¹Cheminformatics and Metabolism, European Bioinformatics Institute (EBI), Cambridge, UK ²Xemistry GmbH, Königstein, Germany ³Faculté de Chimie, Université de Strasbourg, France. ⁴Bayer Business Services GmbH, Leverkusen, Germany.

Email: Christoph Steinbeck^{*} - steinbeck@ebi.ac.uk;

^{*}Corresponding author

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Introduction

The paper will be about

- how to register compounds in RDBMS
- strategies for prescreening
- fingerprinting and solid criteria about how to design fingerprints and what frequency distribution is best
- Alternatives to Fingerprint prescreening
- fastest implementations of substructure queries
- and how to project that onto SQL databases
- Stereochemistry searches
- How to implement wild-card searching
- Notoriously difficult cases, structures where many cartridges on the market fail

Next section

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

A sub section

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia

consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?

Another section

Some sub section

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum hic tenetur a sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat

And also

At vero eos et accusamus et iusto odio dignissimos ducimus qui blanditiis praesentium voluptatum deleniti atque corrupti quos dolores et quas molestias excepturi sint occaecati cupiditate non provident, similique sunt in culpa qui officia deserunt mollitia animi, id est laborum et dolorum fuga. Et harum quidem rerum facilis est et expedita distinctio. Nam libero tempore, cum soluta nobis est eligendi optio cumque nihil impedit quo minus id quod maxime placeat facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum re-

rum hic tenetur a sapiente delectus, ut aut reiciendis voluptatibus maiores alias consequatur aut perferendis doloribus asperiores repellat.

Conclusion

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?

Competing interests

The authors declare that they have no competing interests.

Authors contributions

Bla bla

Acknowledgments

Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur

References

1. Hagadone T, Michael S: **Integrating chemical structures into an extended relational database system.** *Chemical Structures 2: The International Language of Chemistry: Proceedings of the 2nd International Conference* 1995, :257–269.

2. Hagadone T, Schulz M: **Capturing Chemical Structure Information in a Relational Database System: The Chemical Software Component Approach.** *Journal of Chemical Information and Computer Sciences* 1995, **35**(5):879–884.
3. Frome J: **Searching Chemical Structures.** *Journal of Chemical Documentation* 1964, :43–45.
4. Guha R, Howard MT, Hutchison GR, Murray-Rust P, Rzepa H, Steinbeck C, Wegner J, Willighagen EL: **The Blue Obelisk - Interoperability in Chemical Informatics.** *Journal of Chemical Information and Modelling* 2005, **46**(3):991–998.
5. Steinbeck C, Han YQ, Kuhn S, Horlacher O, Luttmann E, Willighagen E: **The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics.** *Journal of chemical information and computer sciences* 2003, **43**(2):493–500.
6. Steinbeck C, Hoppe C, Kuhn S, Guha R, Willighagen EL: **Recent Developments of The Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics.** *Current pharmaceutical design* 2006, **12**(17):2111–2120.
7. Steinbeck C, Krause S, Kuhn S: **NMRShiftDB - Constructing a free chemical information system with open-source components.** *Journal of chemical information and computer sciences* 2003, **43**(6):1733–1739.
8. Steinbeck C, Kuhn S: **NMRShiftDB - compound identification and structure elucidation support through a free community-build web database.** *Phytochemistry* 2004, **65**(19):2711–2717.
9. **The PGChem project.** <http://pgfoundry.org/projects/pgchem/> 2009.
10. **The MyChem project.** <http://mychem.sourceforge.net> 2009.
11. **The OrChem project.** <http://orchem.sourceforge.net> 2009.
12. **The ChemiSQL umbrella project for Open Source chemical search engines.** <http://chemdb.sourceforge.net> 2009.

Figures