МОНИТОРИНГ СТРУКТУРНО-ХИМИЧЕСКИХ БАЗ ДАННЫХ

[1]Круковская Н.В., [2]Зибарева И.В.
[1]Институт органической химии им. Н.Д.Зелинского РАН, г.Москва
[2]Институт органической химии им. Н.Н.Ворожцова СО РАН, г.Новосибирск

[1]Круковская Н.В.
[2]Зибарева И.В.
[1]Институт органической химии им. Н.Д.Зелинского РАН
[2]Институт органической химии им. Н.Н.Ворожцова СО РАН

Задача учета электронных информационных ресурсов является достаточно сложной. Пользователю трудно ориентироваться в обилии слабо структурированной информации, выложенной на сайтах или найденной с помощью поисковых роботов типа Google. Для того чтобы целенаправленно обратиться к первоисточникам научно-технической информации, а именно, к статьям, патентам, диссертациям, используются вторичные источники научно-технической информации: реферативные журналы или их современные аналоги — электронные базы данных. В данном сообщении описаны крупнейшие мировые базы данных по химии и смежным областям, где возможен поиск информации о конкретном соединении и/или классах соединений. Только структурно-химические базы данных (СХ БД) позволяют узнать упомянуто ли данное вещество в статьях, патентах, каталогах. Первые электронные СХ БД возникли еще в 60-70-х годах XX века, более широкое распространение БД связано с развитием теледоступа и приходится на 80-90 годы [1,2]. За это время неоднократно менялось программное обеспечение, производители БД, поэтому не всегда можно понять, что новая информационно-поисковая система состоит в основном из тех же БД, печатные варианты которых хорошо известны, например Chemical Abstracts, Beilstein, Gmelin. Что представляет собой информационное наполнение SciFinderScholar, Web of Knowledge и Discovery Gates? Таблица I полного текста доклада дает ответ на этот вопрос.

С появлением СХ БД возникла новая поисковая возможность, которую трудно реализовать при использовании печатных информационных ресурсов. Это подструктурный поиск — поиск всех химических соединений, содержащих общий фрагмент. Эта функция стала возможна, когда при описании индексируемых веществ начали использовать в качестве дескрипторов графы. С помощью графа каждая связь в молекуле описана как одинарная, двойная, тройная или полуторная (например, в бензоле), различаются связи цепочечные и циклические, каждый атом в молекуле (узел) также опознается как определенный элемент и как цепочечный или циклический [1]. В полном тексте в таблице II доклада приводится детальное описание крупнейших мировых молекулярных СХ БД, во-первых, количество соединений, во-вторых, тип соединений, в-третьих, с какого года охвачена информация в БД, в четвертых, какое количество наименований журналов отражено в БД, в-пятых, какое количество библиографических ссылок присутствует в БД.

Более сложным вариантом СХ БД являются реакционные базы данных, где описаны схемы превращения веществ, как одностадийные, так и многостадийные, с указанием катализаторов, растворителей и выходов на различных стадиях. Кроме того, в большинстве СХ БД помимо данных о соединениях и их химических свойствах, приводятся физико-химические свойства и другая фактографическая информация. Поэтому, в таблице III описывается разделение СХ БД на молекулярные, реакционные, фактографические БД, с указанием какой тип фактографических данных представлен в БД.

Процесс создания СХ БД требует высококвалифицированного труда программистов, инженерного персонала, индексаторов. Поэтому, доступ к СХ БД, как правило, платный, причем СХ БД гораздо более дорогие по сравнению с библиографическими БД. Отсутствие необходимой централизованной материальной поддержки вынуждает российские научные организации, которым необходимы СХ БД, отказываться от их приобретения, что негативно сказывается на их инновационной активности. Что касается центральных научных библиотек, то они более нацелены на полнотекстовые электронные информационные ресурсы. Существующие проблемы доступа должны стать предметом обсуждения химического сообщества, необходимо выработать принципиальную концепцию — кто может вернуть в центральные библиотеки или библиотеки институтов важнейшие информационные ресурсы.

Литература

1. *Blake James E., Farmer Nick A., Haines Reginald C.* An Interactive Computer Graphics System for Processing Chemical Structure Diagrams // Journal of Chemical Information and Computer Sciences. 1977. №4. P.223-228.
2. *Dumont Lise M.* Computerized Structural Databases // Canadian Chemical News. 1988. №6. P.17-20.
3. *Потапов В.М., Кочетова Э.К.* Химическая информация. М.: Химия, 1988.
4. *Wagner A.B*. SciFinder Scholar 2006: An Empirical Analysis of Research Topic Query Processing // J.Chem.Inf.Model. 2006.