

# Поиск химической информации в базах данных сети STN International

*Инна Владимировна Зибарева*

**Et.nsu.ru, дата размещения 1.12.2014**

## **Аннотация**

Курс «Поиск химической информации в базах данных сети STN International» является обязательным для студентов IV курса Факультета естественных наук НИУ-НГУ, обучающихся по программе бакалавриата по направлению подготовки 020100 «Химия».

Цель курса – формирование у студентов профессиональных навыков работы с отечественными и зарубежными базами данных и информационно-поисковыми системами (БД и ИПС) по химии, в том числе патентными.

Курс студентам, специализирующимся по кафедре «Органическая химия», преподается в первом семестре, по кафедре «Аналитическая химия» – во втором семестре. Результатом обучения в рамках курса является итоговая оценка (не дифференцированный зачет).

Рабочая программа курса «Поиск химической информации в научно-технических базах данных» включает в себя обзор основных для профессионального химика современных источников информации – баз данных (БД) и информационно-поисковых систем (ИПС), изучение их особенностей и областей применения, методов поиска различной специализированной информации – библиографической, структурно-химической, фактографической и иной.

Курс предусматривает следующие формы организации учебного процесса: лекции, практические занятия в учебных и реальных БД и ИПС, контрольные работы, самостоятельная работа студента, консультации, сдача зачета.

Рабочая программа курса рассчитана на 72 академических часа – 14 лекционных часов, 3 часа практических занятий, 29 часов прохождения контрольных точек в течение семестра (включая домашние задания), 2 часа зачета, а также 24 часа самостоятельной работы студентов. Результатом обучения в рамках курса является итоговая оценка (не дифференцированный зачет).

Полученные в рамках курса знания позволят студентам выработать навыки самостоятельного поиска и обработки специализированной химической информации с использованием поисково-аналитических возможностей современных БД и ИПС, необходимые для последующей профессиональной деятельности.

## 1. Цели и задачи учебной дисциплины

**Цель курса** «Поиск химической информации в базах данных сети STN International» – формирование у студентов профессиональных навыков работы с отечественными и зарубежными компьютерными информационными ресурсами (БД и ИПС) по химии, в том числе патентными. На лекциях студенты получают основные знания о современных компьютерных информационных ресурсах в области химии, методах поиска специализированной информации в наиболее авторитетных мировых БД и ИПС, подходах к разработке стратегий поиска релевантной информации. На семинарских занятиях – разбирают типовые задачи различной сложности, учатся проводить поиск информации по тематике и веществу в библиографических, структурно-химических, фактографических и иных БД и ИПС. В ходе обучения студенты интенсивно работают со вспомогательной литературой и релевантными информационными ресурсами, доступными по подписке в НИУ-НГУ или институтах СО РАН, а также бесплатно в сети Интернет.

**Результаты освоения курса** студентами – обладание систематизированными знаниями о современных источниках химической информации, владение современными приемами и методами получения релевантной информации, приобретение практических навыков проведения разнообразных поисков в БД и ИПС.

## 2. Содержание учебной дисциплины

Рабочая программа курса «Поиск химической информации в научно-технических базах данных» включает в себя обзор основных для профессионального химика современных источников информации – баз данных (БД) и информационно-поисковых систем (ИПС), изучение их особенностей и областей применения, методов поиска различной специализированной информации – библиографической, структурно-химической, фактографической и иной.

Курс предусматривает следующие формы организации учебного процесса: лекции, практические занятия в учебных и реальных БД и ИПС, контрольные работы, самостоятельная работа студента, консультации, сдача зачета.

Результатом обучения в рамках курса является итоговая оценка (не дифференцированный зачет).

Рабочей программой курса предусмотрены следующие виды контроля: текущий и итоговый. **Текущий контроль** включает контроль посещаемости занятий, сдачу заданий для самостоятельной работы и написание контрольных работ. Для допуска к зачету, студент ходе обучения должен: посетить не менее 50% занятий; выполнить 6 контрольных работ. **Итоговый контроль** включает выполнение зачетного задания, состоящего в составлении поискового запроса и проведении поиска информации в релевантной БД или ИПС по тематике собственной курсовой (дипломной) работы.

Рабочая программа курса рассчитана на 72 академических часа. Общая трудоемкость дисциплины 2 зачетные единицы.

Программой предусмотрены 14 лекционных часов, 3 часа практических занятий, 29 часов прохождения контрольных точек в течение семестра (включая домашние задания), 2 часа зачета, а также 24 часа самостоятельной работы студентов.

Полученные в рамках курса знания позволят студентам выработать навыки самостоятельного поиска и обработки специализированной химической информации с использованием поисково-аналитических возможностей современных БД и ИПС, необходимые для последующей профессиональной деятельности.

### 3. Учебно-методические материалы дисциплины

#### Операторы связи поисковых терминов

<i>Логические операторы</i>	
OR	Объединяет синонимы. Ответы содержат все (любой из) синонимов
AND	Объединяет разные концепты. Ответы содержат все концепты
NOT	Исключение концепта из набора ответов

<i>Операторы близости</i>	
(L)	Связывает термины в одном индексе – одно заглавие, один реферат, один термин индексирования
(S)	Связывает термины в одном предложении – в заглавии, в реферате или в терминах индексирования
(A)	Выстраивает термины рядом в произвольном порядке
(W)	Выстраивает термины рядом в заданном порядке

#### Символы усечения (маскирования)

<i>Символ</i>	<i>Определение</i>	<i>Пример</i>	<i>Будут найдены</i>
?	От 0 до любого числа символов в конце термина	?gene?	abiogenesis partenogenesis osteogenesis generates
#	0 или 1 символ в конце термина	grow##	grow grows growth
!	1 символ внутри или в конце термина	t!!th	teeth tooth truth
		amin!	amine amino

*Символы усечения можно объединять внутри одного термина, разрешено многократное использование символов # и !*

## Основные команды

<i>Команда</i>	<i>Действие</i>	<i>Результат</i>	<i>Пример</i>
FILE (FIL)	Ввод одной или нескольких БД (их кластера) для проведения поиска	Выбор БД для поиска. Получение сведений об информационном охвате и обновлениях БД	=> FILE CAPLUS => FIL PATENTS
EXPAND (E)	Просмотр поисковых терминов в индексе для подтверждения наличия нужного термина в БД; проверки написания термина в БД; идентификации альтернативных форм термина	Алфавитно-цифровой список терминов, соседних заданному термину	=> E STREPTOMYCES
SEARCH (S)	Поиск записей, содержащих термин(ы), и создание набора ответов из этих записей	Создание набора ответов (L#) записей по интересующей теме	=> S CYCLOADDITION
DISPLAY (D)	Вывод на экран результатов поиска	Просмотр результатов из набора ответов в заданном формате	=> D L1 1-2 BIB
LOGOFF (LOG)	Завершение поиска	Окончание работы с сетью STN	=> LOG Y

### Команда DISPLAY (D)

<i>Задание</i>	<i>Вывод по умолчанию</i>	<i>Примечание</i>
L-номер набора ответов	Последний созданный L-номер	Команда D HIS – если нужно уточнить номер набора ответов, созданного ранее
Номер(а) ответа(ов)	Первый ответ	Опции: 1-5 – вывод первых пяти ответов; 1, 5 – просмотр 1-ого и 5-ого ответов
Формат вывода	Библиографическая информация (BIB)	Формат IBIB – библиографическая информация с названиями полей: ABS – реферат; ALL – полная запись

## Основные поисковые индексы базы данных CAPlus

<i>Код</i>	<i>Название индекса</i>	<i>Примечания</i>
TI	<b>T</b> itle	Заглавие публикации
AU	<b>A</b> uthor	Автор
CS	<b>C</b> orporate <b>S</b> ource	Место работы автора
DT	<b>D</b> ocument <b>T</b> ype	Тип документа
LA	<b>L</b> anguage	Язык оригинальной публикации
AB	<b>A</b> Bstract	Реферат
ST	<b>S</b> upplementary <b>T</b> erms	Ключевые слова
IT	<b>I</b> ndex <b>T</b> erms	Концепты, сообщаемые в документе
RL	<b>CAS R</b> oLe	Роль вещества
RE	<b>R</b> Eferences	Ссылки в оригинальной публикации

## Индексы для уточнения набора ответов

<i>Ограничение</i>	<i>Индекс</i>	<i>Пример</i>
Типом документа	/DT	Патенты: => S L7 AND PATENT/DT Статьи из журналов: => S L7 AND JOURNAL/DT
Языком	/LA	Немецкий язык: => S L7 AND GERMAN/LA
Временем	/PY	Год: => S L7 AND 1996/PY или => S L7 AND PY=1996 Период: => S L7 AND 1994-1996/PY или => S L7 AND PY>=2001
Автором	/AU	=> S IVANOV A?/AU
Организацией	/CS	=> S NOVOSIBIRSK UNIVERSITY/CS

## Стратегия поиска информации по ключевым словам в библиографических БД и ИПС

<b>1</b>	<b>Формулировка поискового запроса</b>	
<b>2</b>	<b>Составление поискового предписания</b>	
	– выбор основных концептов и синонимов	
	– выбор логических операторов	AND, OR, NOT
<b>3</b>	<b>Проведение предварительного поиска</b>	
	Ввод БД	=> FILE CA
	Проверка поисковых терминов	=> E ANTIBIOTIC
	Создание набора поисковых терминов	Множественные формы, сокращения и усечения для альтернативных терминов
	Проведение поиска	=> S STREPTOMYCES AND ANTIBIOTIC# AND ANTITUMOR?
<b>4</b>	<b>Оценка ответов с помощью бесплатных команд / идентификация дополнительных терминов</b>	=> D SCAN
<b>5</b>	<b>Уточнение стратегии поиска</b>	
	Учет дополнительных терминов	=> S (ANTITUMOR OR ANTI-TUMOR OR ANTITUMOUR OR ANTICANCER OR NEOPLASM INHIBIT?)
	Применение операторов близости	=> S L2 (S) L3 (S) L4
<b>6</b>	<b>Детальный вывод ответа(ов)</b>	=> D L6 2 IBIB ABS

## Рекомендации для поиска по имени автора

<i>Имя</i>	<i>Пример</i>	<i>Рекомендация для ввода</i>	<i>Пример</i>
Если неизвестно, в какой форме содержится в БД	Karl Wurst Karl A. Wurst K. A. Wurst	Фамилия и инициал	WURST K/AU
Имеет внутреннюю пунктуацию – апострофы, дефисы	O'Brian	Варианты написания с пунктуацией и без	OBRIAN/AU O BRIAN/AU
Содержит внутренние пробелы	La Bar	Варианты написания с пробелом и без	LA BAR/AU LABAR/AU
Содержит умляут	Müller	Варианты с замещениями: ae → ä; oe → ö; ue → ü	MUELLER/AU MULLER/AU
Неясно, что имя, а что фамилия	Chang Cheng	Используя оба слова как фамилию	CHANG/AU CHENG/AU
Транслитерировано, например, с кириллицы	Bagryanski	Используя альтернативное написание	BAGRYANSKII/AU BAGRYANSKY/AU

## Рекомендации для поиска по названию организации

<i>Название</i>	<i>Пример</i>	<i>Рекомендация для ввода</i>	<i>Пример</i>
Изменилось со временем	Corning Glass Works USA → Corning USA	Используя постоянную часть	CORNING/CS
Изменилось после реорганизации (слияния)	Ciba-Geigy + Sandoz → Novartis	Используя новое и старые названия	(CIBA GEIGY OR SANDOZ OR NOVARTIS)/CS
Возможны разные написания	DuPont Du Pont Proctor and Gamble Proctor & Gamble Intel Corp. Intel Corporation	Используя оба варианта Исключение из запроса and или & Исключение Co., Corp., Inc. и др. из запроса	(DUPONT OR DU PONT)/CS PROCTOR GAMBLE/CS INTEL/CS
Различается для подразделений (филиалов)	Rockwell International Science Center и Rockwell International Electron Research Center	Используя общую часть	ROCKWELL/CS

## Роли веществ Chemical Abstracts Service<sup>a</sup>

<b>ANST</b> Analytical study	<b>PREP</b> Preparation <sup>e</sup>
<b>ANT</b> Analyte	<b>BMF</b> Bioindustrial manufacture
<b>AMX</b> Analytical matrix	<b>BPN</b> Biosynthetic preparation
<b>ARG</b> Analytical reagent use	<b>BYP</b> Byproduct
<b>ARU</b> Analytical role, unclassified	<b>CPN</b> Combinatorial preparation <sup>c</sup>
<b>BIOL</b> Biological study	<b>IMF</b> Industrial manufacture
<b>ADV</b> Adverse effect, including toxicity	<b>PUR</b> Purification or recovery
<b>AGR</b> Agricultural use	<b>PNU</b> Preparation, unclassified <sup>f</sup>
<b>BAC</b> Biological activity or effector, except adverse <sup>b</sup>	<b>SPN</b> Synthetic preparation
<b>BCP</b> Biochemical process <sup>c</sup>	<b>PROC</b> Process
<b>BMF</b> Bioindustrial manufacture	<b>BCP</b> Biochemical process <sup>c</sup>
<b>BOC</b> Biological occurrence <sup>b</sup>	<b>BPR</b> Biological process <sup>b</sup>
<b>BPN</b> Biosynthetic preparation <sup>c</sup>	<b>GPR</b> Geological or astronomical process
<b>BPR</b> Biological process <sup>b</sup>	<b>PEP</b> Physical, engineering, or chemical process
<b>BSU</b> Biological study, unclassified	<b>CPS</b> Chemical process <sup>g</sup>
<b>BUU</b> Biological use, unclassified	<b>EPR</b> Engineering process <sup>g</sup>
<b>COS</b> Cosmetic use <sup>c</sup>	<b>PYP</b> Physical process <sup>g</sup>
<b>DGN</b> Diagnostic use <sup>c</sup>	<b>REM</b> Removal or disposal
<b>DMA</b> Drug mechanism of action <sup>c</sup>	<b>PRPH</b> Prophetic substance <sup>h</sup>
<b>FFD</b> Food or feed use	<b>RACT</b> Reactant or reagent <sup>b, g</sup>
<b>MFM</b> Metabolic formation <sup>b</sup>	<b>RCT</b> Reactant <sup>i</sup>
<b>NPO</b> Natural product occurrence <sup>c</sup>	<b>CRT</b> Combinatorial reactant <sup>c</sup>
<b>PAC</b> Pharmacological activity <sup>c</sup>	<b>RGT</b> Reagent <sup>c</sup>
<b>PKT</b> Pharmacokinetics <sup>c</sup>	<b>CRG</b> Combinatorial reagent <sup>c</sup>
<b>THU</b> Therapeutic use	<b>USES</b> Uses
<b>CMBI</b> Combinatorial study <sup>c</sup>	<b>AGR</b> Agricultural use
<b>CPN</b> Combinatorial preparation <sup>c</sup>	<b>ARG</b> Analytical reagent use
<b>CRT</b> Combinatorial reactant <sup>c</sup>	<b>BUU</b> Biological use, unclassified
<b>CRG</b> Combinatorial reagent <sup>c</sup>	<b>CAT</b> Catalyst use
<b>CST</b> Combinatorial study <sup>c</sup>	<b>COS</b> Cosmetic Use <sup>c</sup>
<b>CUS</b> Combinatorial use <sup>c</sup>	<b>CUS</b> Combinatorial use <sup>c</sup>
<b>FORM</b> Formation, nonpreparative	<b>DEV</b> Device Component use <sup>f</sup>
<b>FMU</b> Formation, unclassified	<b>DGN</b> Diagnostic use <sup>c</sup>
<b>GFM</b> Geological or astronomical formation	<b>FFD</b> Food or feed use
<b>MFM</b> Metabolic formation <sup>b</sup>	<b>MOA</b> Modifier or additive use
<b>NANO</b> Nanomaterial <sup>d</sup>	<b>NUU</b> Other use, unclassified <sup>j</sup>
<b>OCCU</b> Occurrence	<b>POF</b> Polymer in formulation
<b>BOC</b> Biological occurrence <sup>b</sup>	<b>TEM</b> Technical or engineered material use
<b>GOC</b> Geological or astronomical occurrence	<b>THU</b> Therapeutic use
<b>NPO</b> Natural product Occurrence <sup>c</sup>	<b>Specific roles</b> that are not associated with any super roles:
<b>OCU</b> Occurrence, unclassified	<b>MSC</b> Miscellaneous
<b>POL</b> Pollutant	<b>PRP</b> Properties

<sup>a</sup> Супер-роли имеют 4-буквенные коды, конкретные роли – 3-буквенные. Под каждой супер-ролью перечислены конкретные роли, которые будут найдены при поиске по супер-роли.

<sup>b</sup> Используется в Chemical Abstracts (CA) с тома № 66 (1967 г.) по том № 135 (2001 г.).

<sup>c</sup> Используется в CA, начиная с тома № 136 (2002 г.).

<sup>d</sup> Используется в CA, начиная с тома № 116 (1992 г.).

<sup>e</sup> Супер-роль PREP добавлена к записям до 1907 г.

<sup>f</sup> Используется в CA с тома № 66 (1967 г.) по том № 145 (2006 г.).

<sup>g</sup> Используется в CA с тома № 136 (2002 г.) по том № 145 (2006 г.).

<sup>h</sup> Используется в CA с 2003 г. по настоящее время.

<sup>i</sup> Поиск по роли RCT находит ссылки в CA с тома № 66 (1967 г.) по настоящее время. Поиск по супер-роли RACT находит ссылки RCT, CRT, RGT или CRG, начиная с тома № 136 (2002 г.).

<sup>j</sup> Начиная с 2002, поисковый текст для роли NUU изменился из Nonbiological use, unclassified/RL в Other use, unclassified/RL. Поиск по роли NUU/RL используется для поиска записей в CA, начиная с тома № 66 (1967 г.) по настоящее время.

## Тезаурус CA Lexicon

### Коды тезауруса CA Lexicon

<i>Код</i>	<i>Описание</i>
ALL	Все релевантные термины, за исключением связанных (LT)
MAX	Все релевантные термины, включая связанные (LT)
BT – Broader Terms	Более широкие термины
HIE – Hierarchy Terms	Термины BT и NT, входящие в иерархию данного термина
HNTE – History Note	Примечание – история введения термина
KT – Keyword Term	Термины, содержащие ключевые слова
LT – Linking term	Связанный термин, модифицирующий информацию из индексного заголовка / index heading modifying information
NEW	Новые термины, заменившие старые (OLD)
NOTE – Notes	Примечания
NT – Narrower Term	Более узкий термин
OLD	Старый термин, замененный новым (NEW)
PFT – Preferred Term	Предпочтительный термин (OLD, NEW, USE, UF)
RT – Related Term	Родственный термин
RTCS – Related Chemical Substance Term	Родственный термин для химического соединения
STD – Standard Term	Стандартный термин (BT, HNTE, Note, NT, RT, RTCS)
UF – Used For	Используемый термин (не предпочтительный синоним)
USE	Термин, который следует использовать

### Фрагмент тезауруса CA Lexicon для катализаторов

<i>№</i>	<i>Количество записей</i>	<i>Код</i>	<i>Термин</i>
1	110818	=>	Catalysts/CT
		HNTE	Valid heading during volume 21 (1927) to present
2	2	OLD	Activator/CT
3	79931	OLD	Catalysts and catalysis/CT
4	53	OLD	Promoter action/CT
5	1606	OLD	Promoters/CT
6	2	OLD	Strengtheners/CT
7		UF	Activators/CT
8		UF	Catalyst/CT
9		UF	Reaction catalysts/CT
10		UF	Supported catalysts/CT
11	65	NT1	Abstraction reaction catalysts/CT
12	2080	NT1	Acylation catalysts/CT
13	1123	NT2	Acetylation catalysts/CT
14	742	NT2	Alkoxyacylation catalysts/CT



## Поиск информации о веществе в структурно-химических БД

### Основные поисковые индексы БД Registry

<i>Код</i>	<i>Поисковый индекс</i>	<i>Примечания</i>
RN	CAS Registry number	Регистрационный номер вещества CAS
CN	CA Index name	Название по номенклатуре CAS
	Other names	Другие названия вещества
CNS	Chemical name segment	Фрагмент химического названия
MF	Molecular formula	Молекулярная формула
LC	Locator code	Другие БД сети STN, имеющие ссылки на данный RN
EPROP	Experimental properties	Экспериментальные свойства вещества
CALC	Calculated properties	Рассчитанные свойства вещества

### Уточнение набора ответов известными свойствами вещества

<i>Свойство</i>	<i>Поиск в индексе</i>	<i>Пример</i>
Показатель кислотности (pKa)	/PKA	=> S PKA<=-0.62 => S -0.62/PKA
Молекулярный вес (Molecular weight)	/MW	=> S MW<200
Свойства «правила пяти» Липинского <sup>a</sup>	/CALC	=> S LIPINSKI/CALC => S LIP/CALC
Температура кипения (Boiling point)	/BP	=> S 150-155/BP => S BP>=150
Плотность (Density)	/DEN	=> S DEN>=1.002
Оптическое вращение (Optical rotatory power)	/ORP	=> S 70-80/ORP
Показатель преломления (Refractive index)	/RI	=> S 1.427/RI => S RI=1.427

<sup>a</sup> Lipinski's Rule-of-Five – эмпирическое правило, по которому лекарственные вещества с хорошей биологической доступностью при пероральном приеме должны обладать следующими характеристиками: число доноров водорода в молекуле (атомов азота или кислорода, связанных с атомами H) не должно превышать 5; число акцепторов водорода (атомов азота или кислорода, не связанных с атомами H) не должно превышать 10; молекулярная масса не должна превышать 500 D; логарифм распределения вещества в системе октанол / вода должен быть не более 5. Название правила связано с тем, что численные границы всех параметров кратны пяти.

## Поиск веществ по химическим названиям

<b>1</b>	<b><i>Поиск регистрационного номера соединения по его названию</i></b>	
	1. Ввод БД Registry	=> FILE REGISTRY
	2. Проверка химического названия	=> E RESVERATROL/CN
	3. Поиск по названию	=> S E3 L1 1 RESVERATROL/CN
	4. Вывод ответов	=> D L1 1 IDE
<b>2</b>	<b><i>Поиск ссылок на соединение</i></b>	
	1. Ввод БД CAPlus	=> FILE CAPLUS => S L1
	2. Поиск по L-номеру из БД Registry	L2 2118 S L1
	3. Оценка найденных ответов	=> D SCAN
<b>3</b>	<b><i>Использование ролей CAS для выделения конкретных записей</i></b> (например, по способам синтеза)	=> S L1/PREP
<b>4</b>	<b><i>Вывод ответов</i></b>	=> D L2 4 IBIB ABS
<b>5</b>	<b><i>Извлечение названий и регистрационных номеров CAS</i></b> из БД Registry для поиска в других БД сети STN International	=> FILE REGISTRY => SEL CHEM L1
<b>6</b>	<b><i>Включение в поиск дополнительных БД сети STN International</i></b>	=> FILE USPATALL PCTFULL => S E1-E2 OR E4-E10
<b>7</b>	<b><i>Удаление дубликатов и вывод ответов на экран компьютера</i></b>	=> DUP REM L2 L7 => D 1 FROM EACH

## Рекомендации по поиску веществ по химическим названиям

<b><i>Название содержит</i></b>	<b><i>Рекомендация</i></b>	<b><i>Пример</i></b>
надстрочные / подстрочные символы, курсив греческие буквы	игнорировать курсив и символы	=> E dichlorometane-d2/CN
апостроф	написать названия букв на латинице, поместив их между точками	=> E .alpha.-acetylnaphthalene/CN
круглые скобки	поместить название в двойные кавычки	=> S "N,N'-dimethyl-1,2-ethanediamine"/CN
квадратные скобки	поместить название в двойные кавычки	=> S "2-(1-acetoxyethyl)furan"/CN
	заменить квадратные скобки круглыми и поместить название в двойные кавычки	=> S "benzo(b)thiophene"/CN

## Поиск веществ по молекулярным формулам

1	Поиск веществ	
1. Ввод БД Registry		=> FILE REGISTRY
2. Проверка присутствия формулы в БД: полных формул – в индексе MF; формул компонент – в Basic Index		=> E C5H13BRN/MF => E C5H13BRN
3. Поиск; возможно уточнение набора ответов известными свойствами или номенклатурным названием		=> S E3 L1 1 C5H13BRN/MF
4. Вывод ответов		=> D L1 1 IDE
2	Нахождение литературных ссылок на найденное вещество	
1. Ввод БД CAPlus		=> FILE CAPLUS
2. Перенос L-номера набора ответов из БД Registry		=> S L1 L2 1071 S L1
3. Оценка ответов		=> D SCAN
3	Добавление кода роли CAS для выделения конкретных записей (например, по способам синтеза)	=> S L1/PREP
4	Вывод ответов для записи в файл	=> D L2 4 IBIB ABS

## Представление формул солей

Обычное	В БД Registry	В индексе MF
$\begin{array}{c} \text{CH}_3-\text{C}=\text{O} \\   \\ \text{O}\cdot\text{Na}^+ \end{array}$	$\begin{array}{c} \text{CH}_3-\text{C}=\text{O} \\   \\ \text{OH}\cdot\text{Na} \end{array}$	C2H4O2.Na
$\begin{array}{c} \text{CH}_3 \\   \\ \text{CH}_3-\text{N}^+-\text{H}\text{Cl}^- \\   \\ \text{CH}_3 \end{array}$	$\begin{array}{c} \text{CH}_3 \\   \\ \text{CH}_3-\text{N}\cdot\text{HCl} \\   \\ \text{CH}_3 \end{array}$	C3H9N.ClH

## Представление формул полимеров

Обычное	В БД Registry	В индексе MF
Гомополимер $\begin{array}{c} \text{H}_2\text{C}=\text{CH} \\   \\ \text{C}_6\text{H}_5 \end{array}$	$\text{H}_2\text{C}=\text{CH}-\text{Ph}$	(C8H8)X
Сополимер винилацетата ( $\text{CH}_3\text{CO}-\text{O}-\text{CH}=\text{CH}_2$ ), винилхлорида ( $\text{CH}_2=\text{CH}-\text{Cl}$ ) и винилфторида ( $\text{CH}_2=\text{CH}-\text{F}$ )	$\begin{array}{l} \text{AcO}-\text{CH}=\text{CH}_2 \\ \text{H}_2\text{C}=\text{CH}-\text{Cl} \\ \text{H}_2\text{C}=\text{CH}-\text{F} \end{array}$	(C4H6O2.C2H3Cl.C2H3F)X

## Поиск семейств веществ

Термин в индексе /MF	Связанные с ним термины в индексе /BI
C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>
C <sub>16</sub> H <sub>32</sub> O <sub>2</sub> .H <sub>3</sub> N	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub> и H <sub>3</sub> N
C <sub>16</sub> H <sub>32</sub> O <sub>2</sub> .Na	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub> и Na

## Поиск веществ по их структурам

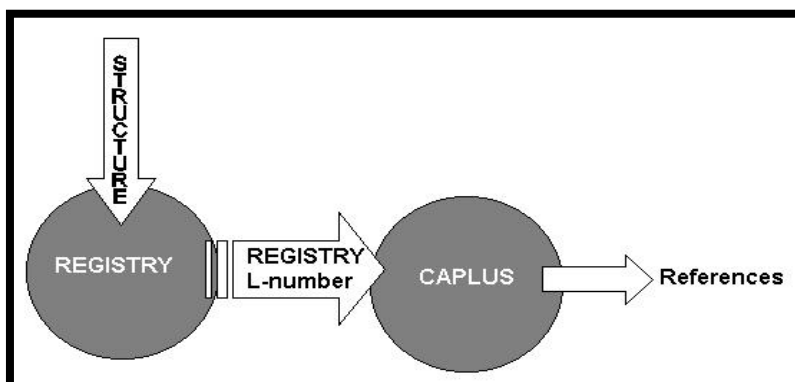
### Цели и типы структурного поиска

<i>Цель</i> – найти:	<i>Тип поиска</i>		
	Exact (EXA)	Family (FAM)	Substructure (SSS)
конкретное вещество	✓	✓	✓
стереоизомеры	✓	✓	✓
изотопно-меченное вещество	✓	✓	✓
соль		✓	✓
смесь		✓	✓
замещенное производное			✓

### Условия и результаты структурного поиска





<i>Условие</i>	<i>Цель</i>
точное соответствие запросу	нахождение синтеза конкретного соединения; выяснение того, было ли вещество получено ранее
близкое соответствие запросу (поиск родственных структур)	биологически активные соли соединения; полимерные материалы; лекарства, содержащие конкретное вещество
структуры, содержащие интересующий скелет или фрагмент	определение аналогов функциональных групп; соотношения структура / свойства (биологическая активность)

### Поиск ссылок на химические структуры



## Стратегия структурного поиска в базе данных Registry

---

<b>1</b>	<b>Рисование структуры:</b> STN Express STN on the Web	
	<b>Сохранение структуры</b>	
<b>2</b>	<b>Вход в сеть STN International</b>	
	Ввод БД Registry	=> FILE REGISTRY
<b>3</b>	<b>Загрузка структуры</b>	
	Проверка загруженной структуры	=> UPLOADED L1 STRUCTURE UPLOADED
		=> D L1 L1 HAS NO ANSWERS L1 STR
<b>4</b>	<b>Проведение пробного поиска</b>	=> S L1 EXACT SAM
	Оценка результатов: вывод структур для оценки; уточнение структуры (при необходимости)	=> D SCAN
<b>5</b>	<b>Проведение поиска по всей БД</b>	=> S L1 EXACT FULL
<b>6</b>	<b>Ввод БД CAPlus</b>	=> FILE CAPLUS
	Поиск ссылок на вещество	=> S L2/PREP => D BIB ABS HITSTR

---

## Поиск по фрагменту структуры

<i>Запрос содержит</i>	<i>Принятое умолчание</i>	<i>Возможности</i>
Атомы с открытыми положениями замещения	Возможно любое замещение в эти положения	Замещение можно заблокировать, например, атомами водорода
Циклические системы	Атрибут циклической системы <i>Isolated / Embedded</i>	Изолировать циклы, чтобы не были найдены конденсированные системы
Цепочки атомов	Атрибут атома в цепи <i>Chain node</i>	Изменить атрибут на <i>Ring / Chain</i> , что позволит находить вещества, содержащие в этом положении атомы, входящие в цикл
Связи в цепочках	Атрибут связи в цепи <i>Chain bond</i>	Изменить характеристику связи на <i>Ring / Chain</i> , что позволит находить вещества с замкнутыми структурами
Насыщенные 6-членные циклы	Характеристика связи <i>Exact / Normalized</i>	Изменить характеристику связи на <i>Exact</i> , если нужны насыщенные циклы; на <i>Normalized</i> , если нужны ароматические циклы
Положения с определенными заместителями	Системные переменные <i>Variables</i>	Изменить атрибуты системных переменных, используя опции <i>Generic Definitions</i> и <i>Element Counts</i>
	Определяемые пользователем <i>G-группы</i>	Определенные пользователем <i>G-группы</i> могут содержать атомы, системные переменные или другие <i>G-группы</i>
Замещение в цикле по одному или нескольким положениям	Возможность задания переменных положений замещения <i>Variable Points of Attachment</i> (VPA)	Опция VPA может использоваться в циклических системах для атомов, системных переменных или <i>G-групп</i>

## Классы соединений в БД данных Registry

<i>Код</i>	<i>Класс веществ</i>
AYS	сплав
CCS	координационные соединения
CTS	зарегистрированные концепты (registered concepts)
GRS	стандартная регистрация
IDS	не полностью определенные соединения
MAN	соединения, зарегистрированные вручную
MNS	минералы
MXS	смеси
PMS	полимеры
RIS	ион-радикалы
RPS	архетипы циклических систем
TIS	табличный состав для неорганических соединений
UVCB	соединения неизвестного или переменного состава или соединения биологического происхождения

### Особенности поиска неорганических соединений в БД Registry

Различаются следующие типы неорганических веществ: *координационные соединения, металлсодержащие органические соли, металлоорганические соединения, металлы, сплавы, и минералы.*

**Координационные соединения** – нейтральные молекулы или ионы, в которых центральный атом (обычно атом металла) связан с другими, причем количество связей не равно валентности центрального атома.

**Металлоорганические соединения** содержат хотя бы один атом углерода, непосредственно связанный с атомом металла.

**Металлы** – элементы, отдающие электроны с образованием положительных ионов (катионов) и в конденсированном состоянии имеющие металлические связи между атомами.

**Сплавы** – смеси металлов с другими металлами, газами или неметаллическими соединениями, образующиеся при расплавлении и не разделяющиеся на компоненты при охлаждении.

**Минералы** – образовавшиеся в природе химические элементы или их соединения, имеющие определенный химический состав и, обычно, характерную форму кристаллов.

Для поиска ссылок на неорганические соединения используются символы элементов (/ELC), при необходимости – число компонент (/NC); для поиска групп элементов – поле Периодических групп (/PG); для поиска сплавов – поля Состав (/MAC) и Относительный состав (/RC).

## Коды периодических групп

Коды периодических групп могут использоваться для поиска семейств элементов в периодах и группах Периодической таблицы.

Они генерируются для всех элементов молекулярной формулы, за исключением углерода и водорода.

Коды семейства элементов хранятся в поле /PG.

												<b>A8</b>									
												2									
<b>A1 A2</b>												He									
3	4											5									
Li	Be											B									
												7									
												N									
												8									
												O									
												9									
												F									
												10									
												Ne									
												13									
												Al									
												14									
												Si									
												15									
												P									
												16									
												S									
												17									
												Cl									
												18									
												Ar									
<b>A1</b>	<b>A2</b>	<b>T1</b> →	<b>B3</b>	<b>B4</b>	<b>B5</b>	<b>B6</b>	<b>B7</b>	<b>B8</b>		<b>B1</b>	<b>B2</b>										
11	12	→	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36			
Na	Mg		Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr			
19	20	→	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54			
K	Ca		Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe			
37	38	→	57	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86			
Rb	Sr		La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn			
55	56	→	89																		
Cs	Rb		Ac																		
87	88																				
Fr	Ra																				

<b>LNTN</b> →	58	59	60	61	62	63	64	65	66	67	68	69	70	71
	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
<b>ACTN</b> →	90	91	92	93	94	95	96	97	98	99	100	101	102	103
	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
<b>SHEL</b> →	104													

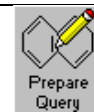


## Поиск по химическим реакциям в базе данных CASREACT

### Стратегия поиска по реакциям

#### 1 Создание структурного запроса по реакции

STN Express  
STN on the Web



Задание направления реакции



Задание роли структурного фрагмента



Спецификация связи, изменяющейся (остающейся неизменной) в реакции



Задание соответствия между атомами реагента и продукта (мэпирование)



Сохранение запроса по реакции



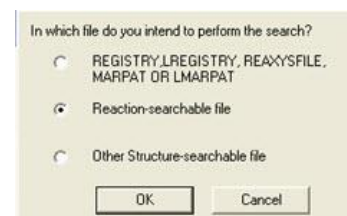
#### 2 Вход (LOGON) в сеть STN



Ввод базы данных CASREACT

=> FILE CASREACT

#### 3 Загрузка запроса в файл реакций сети STN



Проверка загруженной реакции

=> D L1  
L1 HAS NO ANSWERS  
L1 STR

#### 4 Пробный поиск

Оценка результатов:  
вывод реакции для оценки  
уточнение реакции (при необходимости)

=> S L1 SAM

=> D SCAN

=> S L3 (L) NS = 1

#### 5 Полный поиск

=> S L1 FULL

## Термины функциональных групп

Acetal	Halohydrin	Phosphite
Acetyl	Hemiacetal	Phosphonate
Acid halide	Heterocycles	Phosponium
Acyclic alkene	Hydrazide	Phosphorus ylide
Acyclic ketone	Hydrazine	Pi-alkene
Acylmetal	Hydrazone	Pi-alkyne
Alcohols	Hydroperoxide	Pi-allyl
Aldehyde	Hydroxylamine	Primary alcohol
Alkenes	Imide	Primary amine
Alkyl halide	Imine	Purine
Alkyne	Imino ether	Quaternary ammonium
Alkynes	Isocyanate	S-O group
Allene	Isonitrile	Se group
Allyl alcohol	Isothiocyanate	Secondary alcohol
Allyl halide	Ketal	Secondary amine
Amide	Ketene	Selenide
Amidine	Ketenimine	Selenol
Amine oxide	Ketones	Silyl
Amines	Lactam	Silyl enol ether
Anhydride	Lactone	Sulfenyl halide
Aryl halide	Mesyl	Sulfide
Arylsulfonyl	Metal arene	Sulfinate
Azide enol	Metal carbene	Sulfinyl halide
Azine	Metal carbonyl	Sulfonamide
Aziridine	Metal cyclopentadienyl	Sulfone
Azo	Metal halide	Sulfonyl halide
Azoxy	Metal hydride	Sulfonyloxy
Carbamate	Metal metal bond	Sulfoxide
Carbonate	Metal nitrogen	Sulfur ylide
Carbonate derivatives	Metal nitrosyl	Te group
Carboxy derivatives	Metal phosphine	Tertiary alcohol
Carboxylate	Metal sulfur	Tertiary amine
Carboxylic	Metallocarbocycle	Thioacetal
Cephem	Mu-carbonyl	Thioamide
Chloramine	Nitrile	Thiocarbonyl
Cyanamide	Nitrile oxide	Thiocarboxy
Cyanate	Nitrite	Thiocyanate
Cyanohydrin	Nitro	Thioketal
Cyclic alcohol	Nitro	Thiol
Cyclic alkene	Nitro	Thione
Cyclic ketone	Nitrosamine	Thiophenol
Cyclopropyl	Nitroso	Thiourea
Diazo	Nitroxide	Triazene
Diazonium	Null	Trihalide
Diene	O-quinone	Unstd acid
Diimide	Organometal	Unstd aldehyde
Disulfide	Organometallics	Unstd amide
Enamine	Ortho ester	Unstd ester
Enol	Oxime	Unstd ketone
Enol ether	Oxonium	Unstd nitrile
Enyne	P-N group	Unsaturated acid
Episulfide	P-O group	Unsaturated aldehyde
Epoxide	P-quinone	Unsaturated amide
Ether	P-S group	Unsaturated ester
Gem-dihalide	Penam	Unsaturated ketone
Glycol	Peroxide	Unsaturated nitrile
Guanidine	Peroxy acid	Urea
Halides	Peroxy	Vic-dihalide
Haloformate	Phenol	Vinyl halide
	Phosphate	

**Термины классов и относящиеся к ним термины функциональных групп**

Термины классов		Термины функциональных групп
<b>Alcohols</b>	Allyl alcohol	Hemiacetal
	Cyanohydrin	Hydroxylamine
	Cyclic alcohol	Phenol
	Enol	Primary alcohol
	Glycol	Secondary alcohol
	Halohydrin	Tertiary alcohol
<b>Alkenes</b>	Acyclic alkene	Cyclic alkene
	<b>Alkynes</b>	Alkyne
<b>Amines</b>		Enyne
	Amine oxide	Hydroxylamine
	Aziridine	Imine
	Chloramine	Primary amine
	Cyanamide	Secondary amine
	Enamine	Tertiary amine
<b>Carbonate derivatives</b>	Carbamate	Haloformate
	Carbonate	Thiourea
	Guanidine	Urea
<b>Carboxy derivatives</b>	Acid halide	Imide
	Amide	Lactam
	Amidine	Lactone
	Anhydride	Peroxy acid
	Carboxylate	Peroxy ester
	Carboxylic	Thioamide
	Haloformate	Thiocarboxy
	Acid halide	Metal halide
	Alkyl halide	Sulfenyl halide
	Allyl halide	Sulfinyl halide
	Aryl halide	Sulfonyl halide
	Chloramine	Trihalide
<b>Halides</b>	Gem-dihalide	Vic-dihalide
	Haloformate	Vinyl halide
	1,2-C3N2	1,4-C4NO
	1,2-C3NO	1,4-C4NS
	1,2-C3NS	1,4-C4O2
	1,2-C3O2	1,4-C4OS
	1,2-C3OS	1,4-C4S2
	1,2-C3S2	1,4-C5N2
	1,2-C4N2	C2S
	1,2-C4NO	C3N
	1,2-C4NS	C3O
	1,2-C4O2	C3S
	1,2-C4OS	C4N
	1,2-C4S2	C4O
	1,3-C3N2	C4S
	1,3-C3NO	C5N
	1,3-C3NS	C5O
	1,3-C3O2	C5S
	1,3-C3OS	C6N
	1,3-C3S2	C6O
	1,3-C4N2	C6S
	1,3-C4NO	Aziridine
	1,3-C4NS	Cephem
	1,3-C4O2	Episulfide
	1,3-C4OS	Epoxide
	1,3-C4S2	Penam
	1,4-C4N2	Purine
	<b>Ketones</b>	Acyclic ketone
Cyclic ketone		P-quinone
<b>Organometallics</b>		Acylmetal
	Metal arene	Metal phosphine
	Metal carbene	Metal sulfur Metallocarbocycle
	Metal carbonyl	Mu-carbonyl
	Metal cyclopentadienyl	Organometal
	Metal halide	Pi-alkene
	Metal hydride	Pi-alkyne
	Metal metal bond	Pi-allyl
	Metal nitrogen	

## Уточнение результатов поиска по реакции

<i>Уточнение результатов поиска критерию</i>	<i>Используемый индекс</i>
Выход реакции	/YD
Количество стадий	/NS
Регистрационный номер CAS конкретного участника реакции:	
Растворитель	/SOL
Катализатор	/CAT
Реагент	/RGT
Реактант	/RCT
Реагент или реактант	/RRT
Продукт	/PRO
Не продукт	/NPRO
Термин из примечания к реакции (Note)	/BI

## Примеры использования ролей веществ для уточнения набора ответов

<i>Задача</i>	<i>Действия</i>
ограничить найденный набор реакциями, в которых в качестве катализатора используется палладий (регистрационный номер CAS 7440-05-3)	=> S L1 (L) 7440-05-3/CAT
ограничить найденный набор каталитическими реакциями	=> S L1 (L) ANY/CAT
ограничить найденные реакции теми, в которых уксусная кислота (регистрационный номер CAS 64-19-7) – исходное вещество	=> S L1 (L) 64-19-7/NPRO
удалить реакции, в которых в качестве реактанта или реагента используется уксусный ангидрид (регистрационный номер CAS 108-24-7)	=> S L1 (NOTL) 108-24-7/RRT

## Форматы вывода информации в базе данных CASREACT

<i>Формат вывода</i>	<i>Результат вывода</i>
OCC	Количество найденных (HIT) реакций в каждом ответе
CRD	Все найденные (HIT) реакции для каждого ответа в компактной форме
FCRDREF	Первая найденная (HIT) реакция для каждого ответа в компактной форме вместе со ссылкой на источник
CRDREF	Все найденные (HIT) реакции для каждого ответа в компактной форме вместе со ссылкой на источник
FHIT	Первая найденная (HIT) реакция для каждого ответа в полной форме (карта, диаграмма и резюме, включая регистрационный номер CAS для каждого участника реакции)
HIT	Все найденные (HIT) реакции для каждого ответа в полной форме (карта, диаграмма и резюме, включая регистрационный номер CAS для каждого участника реакции)
BIB	Библиографическая информация
ABS	Реферат Chemical Abstracts

#### 4. Контроль изучения дисциплины

**Формой текущего контроля** при прохождении курса «Поиск химической информации в научно-технических базах данных» является контроль посещаемости занятий, сдача заданий для самостоятельной работы и выполнение контрольных работ.

Для допуска к зачету, студент должен: посетить не менее 50% занятий; выполнить не менее 60% заданий для самостоятельной работы и все контрольные работы. При наличии уважительных причин выполнение заданий и (или) контрольных работ может быть перенесено на другой срок в пределах семестра.

<b>Задание</b>	<b>Тема</b>
<b>Контрольная работа № 1</b>	<i>Поиск информации по автору в учебных БД</i>
<b>Контрольная работа № 2</b>	<i>Поиск информации по названию организации в учебных БД</i>
<b>Контрольная работа № 3</b>	<i>Поиск информации по ключевым словам в учебных БД</i>
<b>Контрольная работа № 4</b>	<i>Поиск веществ по их молекулярным формулам в учебных БД</i>
<b>Контрольная работа № 5</b>	<i>Поиск веществ по их названиям в учебных БД</i>
<b>Контрольная работа № 6</b>	<i>Поиск веществ по их структурам в учебных БД</i>
<b>Контрольная работа № 7</b>	<i>Поиск специализированной информации в релевантных БД</i>
<b>Контрольная работа № 8</b> (зачетное задание)	<i>Поиск информации по теме научной (курсовой, дипломной) работы в релевантных БД</i>

**Учебно-методическое обеспечение самостоятельной работы** состоит в том, что задания выдаются студентам в виде печатных материалов и (или) компьютерных файлов. Для выполнения полученных заданий студенты могут использовать релевантные БД и ИПС, доступные в НИУ-НГУ и институтах СО РАН, а также в сети Интернет. Указанная ниже рекомендованная литература и другие материалы доступны в НИУ-НГУ, химических институтах СО РАН и сети Интернет.

#### 5. Литература

1. Зибарева И.В. Химические базы данных сети STN International // Известия АН. Сер. хим. 2012. № 3. С. 679-716.
2. Ridley D.D. Information Retrieval: SciFinder. Wiley, 2009. 214 pp.
3. Хуторецкий В.М. Общие представления о поиске научно-технической информации в режиме онлайн. Базы данных STN International в теледоступе. М: РХТУ, 2000. 42 с.
4. Потапов В.М., Розенман М.И., Кочетова Э.К., Покровский Б.И. Поиск химической информации. Справочное руководство по использованию традиционных и компьютерных средств. М: Изд-во МГУ, 1990. 174 с.