

Што такое Інтэрнэт

Мы заканчваем знаёміць нашых чытачоў з камп'ютэрнай сеткай Інтэрнэт і прапануем наступны артыкул з серыі матэрыялаў, прысвечаных тым, хто збіраецца працаваць у гэтай галіне.

VII. Пошук і прыём інфармацыі ў рэжыме Search

28. Як каротка ахарактарызаваць сутнасць працы ў рэжыме Search?

Існуюць два спосабы здабывання камп'ютэрнай інфармацыі: *Browse* і *Search*.

артыкуле.

Запіс, у сваю чаргу, падзяляецца на шэраг палёў (*fields*). Што да бібліяграфічных баз дадзеных, то найчасцей па асобных палях размяркоўваюцца наступныя порцыі інфармацыі: прозвішчы аўтараў, назва артыкула, назва часопіса, год выдання, рэферат.

Такая структураванасць дазваляе эфектыўна вырашаць даволі разнастайныя праблемы, напрыклад: "Знайсі запісы, у якіх сустракаецца слова XXX", "Знайсі запісы, у якіх сло-

словам *search* (напрыклад, *Search the database...*).

Праілюструем сказанае на прыкладзе базы дадзеных *ESToC*.

У верхняй частцы Галоўнай старонкі *ESToC* (<http://www.elsevier.nl/estoc/>) знаходзяцца навігацыйныя кнопкі; трэцяя ў шэрагу мае назву *Search* і прызначаецца для пераходу ў найменны рэжым працы.

Калі націснуць на навігацыйную кнопку *Search*, на экран выводзіцца пошукавы бланк. *ESToC* можа прапанаваць бланкі трох ступеняў складанасці; найбольш просты з іх (*Basic*) прыведзены на малюнку 1.

31. Якія дзеянні трэба выканаць, каб пачаць працу пошуку?

Пачынаючы пошук, трэба запоўніць пошукавы

некалькіх параметрах, таму на практыцы часцей карыстаюцца мнагаслоўнымі пошукавымі фразамі — гэта значна павялічвае эфектыўнасць працы. Вось толькі правілы складання такіх фразаў могуць адрознівацца ў розных базах дадзеных — адзінага стандарту для пошукавых праграм у свеце няма. Што тут можна парамаганьне — не лянуцца чытаць англійскую дапаможную дакументацыю (гл., напрыклад, гіперсувязь *Quick Reference* на мал.1). Мы ж цяпер коратка ахарактарызуем некаторыя дэталі, на якія трэба звяртаць асаблівую ўвагу.

Пакуль што большая частка пошукавых праграм арыентуецца на заданні, сфармуляваныя з дапамогай апэратараў алгебраічнай логікі (*Boolean search*). Стандартны мінімум апэратараў, якія адназначна ўспрымаюцца ўсімі такімі праграмамі — гэта *AND*, *OR* і *NOT*. Прыядзем прыклады іх выкарыстання ў пошукавых фразах:

aaa AND bbb — знайсці запісы, у якіх ёсць і слова *aaa*, і слова *bbb*;
aaa OR bbb — знайсці запісы, у якіх ёсць або слова *aaa*, або слова *bbb*;
aaa NOT bbb — знайсці запісы, у якіх ёсць слова *aaa*, але няма слова *bbb*.

Іншыя апэратары дазваляюць пабудоваць больш рафінаванае заданне, але трэба мець на ўвазе, што іхсэнс можа вар'іравацца ў розных ба-

аб'ядноўваць аднакарэнныя словы ў адну групу, але вынікі могуць быць і непажаданымі. Напрыклад, *chlor** азначае і блізкія *chlorine, chloride, chlorate*, і далёкі па сэнсу, затое сугучны *chlorophyll*.

2) Даволі распаўсюджаны праграмы, якія ў т.зв. рэжыме *stemming* аўтаматычна ўлічваюць існаванне розных граматычных формаў (слова *run* адпавядае групе *run, runs, running* і г.д.).

3) Шаблон (*wildcard*) замяняе пераменную літару ў сярэдзіне слова (напрыклад, *analy?e* — тое ж, што *analyze* і *analyse*).

Па-рознаму задаецца тэма поля, у якім плануецца праводзіць пошук. Адныя праграмы патрабуюць, каб адпаведныя ўказальнікі былі ўключаныя ў сам тэкст пошукавай фразы. Іншыя прапануюць для гэтага спецыяльны пошукавы бланк з наборам уключальнікаў і/ці меню. На мал.2 прыведзены прыклад такога бланка з заданнем "Знайсі артыкулы аб фасфатах цырконію, апублікаваныя ў 1996 годзе аўтарам *Clearfield 'am* (*ESToC, Combined Search*).

Што датычыць прыцыпаў выбару матэрыялу з базы дадзеных, то тут таксама ёсць два падыходы: (1) строгае падпарадкаванне інструкцыі, закладзенай у пошукавай фразе; (2) трактоўка пошукавай фразы як пажадання, а не загаду. Праграмы апошняга тыпу колькасца па многіх параметрах ацэньваюць адпаведнасць (*relevancy*) знойдзенай інфармацыі заданню, і чытачу перадаецца пакет запісаў, адсартаваных ад больш адпаведнага да менш адпавед-

ваць у фармаце *.txt*.

34. Як можна знайсці URL дакумента патрэбнай тэматыкі?

Web-прастора пастаянна сканіруецца спецыяльнымі робатамі, і ў выніку ствараюцца базы дадзеных — каталогі Web-старонак. Калі далучыцца да такой базы дадзеных, то з дапамогай адпаведнага пошукавага сродку (*search engine*) можна атрымаць URL і нават гіперсувязь да патрэбнага дакумента. Вядома больш дзесятка вялікіх каталогаў; кожны мае свае плюсы і мінусы, звязаныя з аб'ёмам, тэматыкай, глыбінёй індэксавання і г.д. На першым этапе працы мне здаецца метаэгодным азнаямленне з двума пошукавымі сродкамі: *Yahoo!* і *Infoseek*.

Yahoo! (<http://www.yahoo.com/>): 200 тыс. запісаў, размешчаных па тэматычных каталогах і падкаталогах. Падтрымліваюцца рэжымы *Browse* і *Search*. У гэтай базе дадзеных прывабліваюць кароткія (двух-, трохрадковыя), але інфарматыўныя анатацыі да Web-дакументаў.

Infoseek (<http://www.infoseek.com/>): 50 млн. запісаў. Рэжым *Browse* падтрымліваецца, але асноўныя вартасці базы дадзеных сканцэнтраваныя ў выдатнай пошукавай праграме, якая працуе ў рэжымах *Boolean search* і *free-text search*.

Дэталыны разгляд метады працы з *search engines* — тэма асобнай размовы; агульная ж схема пошуку выглядае наступным чынам:

1) Вызваць Web-старонку з пошукавым блан-

Search Facilities	
Basic Search Advanced Search Combined Search	Гіперсувязі да іншых бланкаў
<input type="text"/>	Поле для запісу пошукавай фразы
<input type="button" value="Search"/>	Кнопка "Пачаць пошук"
Quick Reference About the Search Engine	Уключальнікі для выбару параметраў пошуку
<input type="radio"/> Search All scientific fields <input checked="" type="radio"/> Select one or more scientific fields below: <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Chemistry and Chemical Engineering <input type="checkbox"/> Clinical Medicine <input type="checkbox"/> Computer Science <input type="checkbox"/> Earth and Planetary Sciences <input type="checkbox"/> Economics, Business and Management Science <input type="checkbox"/> Engineering, Energy and Technology 	<input type="checkbox"/> Physics and Astronomy <input type="checkbox"/> Social Sciences <input type="checkbox"/> Multidiscipline

Environmental Science and Technology

Search in:

Elsevier Science Catalogue

Elsevier Science Tables of Contents (ESToC)

Max. number of results: 20

Мал. 1

Web-старонкі, як і старонкі звычайнай кнігі, можна гартаць (*browse*) — г.зн., паслядоўна пераходзіць ад аднаго дакумента да другога па сістэме гіперсувязяў або задаючы іх адрасы (*URL*). Методыка працы ў рэжыме *Browse* абмяркоўвалася ў папярэднім артыкуле пра Інтэрнэт.

Search — гэта мэтадыкіраваны пошук Web-дакументаў па пэўных іх прыметах, часцей за ўсё па наяўнасці (ці адсутнасці) у дакументах нейкага слова (ці набору словаў).

Сам браўзер для такой працы не прызначаны. Ён толькі можа злучыць чытача са спецыяльнымі пошукавымі праграмамі (*search and retrieval software; search tool, search engine*), якія дзейнічаюць у межах адной ці некалькіх электронных баз дадзеных.

29. З якіх структурных элементаў складаецца электронная база дадзеных?

Асноўным структурным элементам электроннай базы дадзеных з'яўляецца запіс (*record*). Напрыклад, ва ўжо знаёмым нам масіве *ESToC* ("Змест часопісаў выдавецтва *Elsevier Science*") кожны з запісаў утрымлівае ў сабе звесткі толькі аб адным навуковым

ва *XXX* сустракаецца толькі ў назве артыкула", "Знайсці публікацыі аўтара *YYY*", "Знайсці публікацыі аўтара *YYY*, выдадзеныя ў часопісе *ZZZ* у *NNN* годзе" і г.д.

Зразумела, што (1) для пошуку неабходна ведаць *URL* сервера, на якім знаходзіцца патрэбная база дадзеных; (2) на гэтым серверы павінен быць не толькі інфармацыйны масіў, але і свая пошукавая праграма.

30. Якія прыкметы паказваюць на тое, што з базай дадзеных можна працаваць у рэжыме *Search*?

Знешнім атрыбутам пошукавай праграмы з'яўляецца пошукавы бланк, які мае па меншай меры два элементы: месца (*поле*) для запісу пошукавай фразы і кнопку *Search* (варыянты: *Seek, Find* і г.д.).

Электронныя базы дадзеных, арыентаваныя толькі (ці ў асноўным) на рэжым *Search*, звычайна ўтрымліваюць пошукавы бланк на сваёй Галоўнай старонцы. У іншых выпадках для пераходу ў рэжым *Search* (канешне, калі ён дазваляецца) трэба скарыстацца яшчэ адной, навігацыйнай, кнопкай *Search* або гіперсувязю з

бланк: (1) вызначыць крытэрыі, па якіх павінен адбывацца адбор інфармацыі — запісаць пошукавую фразу (гл. ніжэй); (2) калі на бланку ёсць іншыя палі з пераменнымі параметрамі, адпаведным чынам запоўніць іх; (3) даць загад праграме пачаць пошук — націснуць кнопку *Search*.

(ZIRCONIUM OR Zr) AND PHOSPHATE*	Article title	AND
1996	Year	AND
CLEARFIELD	Editors / Authors	
Search		

Пошукавая фраза

Выбар поля

Выбар аператара

Мал. 2

32. Як фармулюецца пошукавая фраза і якія асаблівасці могуць мець пошукавыя праграмы?

У найбольш простым выпадку пошукавая фраза можа складацца з аднаго слова. Вынікам дзейнасці пошукавай праграмы па такому заданню быў бы выбар з базы дадзеных і вывад на экран усіх запісаў, у якіх сустракаецца гэтае слова.

Асноўнай перавагай камп'ютэрызаванай апрацоўкі інфармацыі з'яўляецца магчымасць хуткага адбору матэрыялу адразу па

зах дадзеных. У прыватнасці, для пошуку ў *ESToC* выкарыстоўваюцца: *ADJ* (аб'яднаня аператарам тэрміны павінны быць у непасрэдным суседстве і менавіта ў такім жа парадку); *W/n* (тэрміны павінны быць у межах *n* слоў адзін ад аднаго і менавіта ў такім жа парадку); *NEAR/n* (тэрміны павінны быць у межах *n* слоў адзін ад аднаго і ў любым парадку).

Апошнім часам у многіх базах дадзеных пошукавую фразу прапануецца запісаць ўвогуле без аператараў. Перш чым рабіць гэта, разбярыцеся, як праграма інтэрпрэтуе такі

нага. Вышэй адзначалася, што пошукавая фраза з аператарамі — не адзіны магчымы спосаб фармулявання задання. Альтэрнатыўны метад, для якога дастаткова тэкста, запісанага на натуральнай мове, — *free-text search* — усё часцей выкарыстоўваецца ў *WWW*. У многіх базах дадзеных чытач сам можа выбраць рэжым пошуку — *Boolean search* ці *free-text search*.

33. Ці ёсць якія-небудзь асаблівасці ў метадыцы захоўвання інфармацыі, атрыманай у рэжыме

тэкст. Справа ў тым, што замест прапушчаных аператараў адныя сістэмы аўтаматычна падстаўляюць *AND*, другія — *ADJ*, а трэція — *OR*. Тут хаваецца вялікая верагоднасць непаразумення паміж чытачом і аўтарам праграмы.

Розныя спосабы існуюць для таго, каб пошукавую фразу для многаварыянтнага задання зрабіць кампактнай:

1) Функцыя скарачэння (*truncation*) замяняе частку слова ўмоўным сімвалам (звычайна, зорчкай *). Гэта дазваляе

Search?

Незалежна ад рэжыму працы для захоўвання інфармацыі выкарыстоўваецца адна і тая ж метадыка (падрабязнасці — у мінулым артыкуле пра Інтэрнэт).

Мне давялося сутыкнуцца толькі з адной асаблівасцю — падчас запісу на дыск файлаў з расшыраннем *.cdg* захоўваецца толькі пошукавы бланк, а знойдзеная інфармацыя можа згубіцца. У такіх выпадках трэба скапіраваць тэкст у кліпбарт, перанесці яго ў тэкставы рэдактар і запіс-

кам (напрыклад, <http://www.yahoo.com>).

2) У пошукавым полі запісаць набор тэрмінаў, якія, на вашу думку, абавязкова павінны быць у патрэбным Web-дакуменце (асабліва ў яго загаловку). Націснуць кнопку *Search* (ці *Find*, ці *Seek*).

3) На экран выводзіцца спіс — *URL* і першыя радкі з тых дакументаў, якія ў большай ступені адпавядаюць вашаму заданню. Выбраць патрэбны дакумент і шчоўкнуць па гіперсувязі — *тэксту URL*. Камп'ютар злучыць вас з патрэбнай Web-старонкай.

Поспехаў ВАМ у пошукавай працы!

А. РАГОЙША
Сорасаўскі вучэбна-інфармацыйны цэнтр па хімічнай адукацыі
Кафедра агульнай хіміі і метадыкі выкладання хіміі