

Analýza možností využití automatizovaných systémů vědeckotechnických informací pro rozvoj biotechnologií

663.18 579

Ing. RICHARD BUREŠ, CSc., Mikrobiologický ústav ČSAV, Praha

Klíčová slova: automatizované systémy vědeckotechnických informací, biotechnologie, magnetopáskové báze dat, referátorové informační služby, selektivní distribuce informací, on-line systémy zpracování informací.

Úvod

Oblast biotechnologických procesů je možno rozdělit do dvou základních kategorií. Zaprvé jsou to tradiční biotechnologické výroby spojené s klasickými fermentačními postupy, mezi které patří produkce piva, vína, výroba krmných kvasnic atp. Do druhé kategorie se řadí biotechnologické postupy vycházející z nových poznatků v oblasti mikrobiologie, genového inženýrství, procesního inženýrství a další oborů. Tímto rozdělením se bude řídit i zde popisovaný systém informačního zajištění biotechnologií.

Tradiční biotechnologické výroby jsou poměrně dobře definovanou oblastí, ve které lze dosti přesně vymezit okruh požadovaných informací i primární (časopisecké) a sekundární (referátové) zdroje, které tyto informace obsahují. Z referátových služeb, které jsou pro tyto účely zvláště vhodné, lze jmenovat například Chemical Abstracts, Biological Abstracts, Food, Science and Technology Abstracts (FSTA). Všechny tyto a jim podobné referátové časopisy existují již několik desítek let (např. Chemical Abstracts od roku 1907) a mají svůj spolehlivý, ustálený a dobře fungující systém abstrahování několika tisíc časopisů z daného (event. i příbuzných) oborů. Uživatel takovýchto informačních systémů po získání jistě zkušenosti nemá zpravidla problémy s jejich využíváním.

Poněkud jiná situace nastává v případě potřeby sledovat dílčí způsobem nebo globálně nové biotechnologické procesy a perspektivy jejich budoucích aplikací. Především v oblasti komercializace biotechnologických výrob se důležité informace začaly objevovat teprve před několika lety a dosud si nenašly své pevné místo ani v oblasti primárních zdrojů (průběžně vznikají a opět zanikají nové specializované časopisy) ani v oblasti sekundárních zdrojů (tradiční referátové služby se snaží tuto problematiku postupně včlenit pokud možno integrálně do svého informačního programu a vedle toho vznikají nové referátové služby specializované na oblast výzkumu, výroby, a prodeje biotechnologicky vyráběných produktů). Mezi tyto nově vznikající sekundární informační zdroje patří např.: BIOTECHNOLOGY Abstracts, společnosti Derwent Pub., Ltd nebo Current BIOTECHNOLOGY Abstracts, produkované Royal Society of Chemistry, University of Nottingham, UK a některé další.

V tomto příspěvku budou stručně popsány a zhodnoceny obě jmenované kategorie informačních zdrojů, a to jak z hlediska průběžného poskytování požadovaných informací vědeckovýzkumné základě, tak také z hle-

diska využitelnosti pro analýzy vývojových trendů v této oblasti. Pozornost bude soustředěna především na využívání systémů automatizovaného zpracování vědeckotechnických informací za pomoci výpočetní techniky.

Klasické referátové služby

Do této kategorie budeme řadit takové referátové služby, které existují již řadu let, sledují rozsáhlý a dobře definovaný okruh časopisů z určitého oboru i oborů nějakým způsobem příbuzných a používají pro anotování vybraných článků z časopisů, konferenčních sborníků, disertačních prací, patentů, knih aj. ustálený okruh deskriptorů (klíčových slov). Zpravidla je v takovém případě k dispozici tezaurus, neboli slovník závazných předmětových hesel, kterých se používá k popisu věcného obsahu primárního anotovaného dokumentu. Jako příklady takovýchto referátových služeb je možno jmenovat: CHEMICAL Abstracts (anotují zcela nebo částečně okolo 14 tisíc chemicky zaměřených periodik), BIOLOGICAL Abstracts, EXCERPTA Medica, Food, Science and Technology Abstracts.

Pro zefektivnění práce při vyhledávání požadovaných informací se využívá automatizovaného zpracování informačních záznamů, které svou formou i obsahem jsou ekvivalentní záznamům v tištěných referátových časopisech, avšak jsou uloženy na magnetických páskách a v této formě předávány jednotlivým uživatelským organizacím. Ty je pak zpracovávají pomocí speciálních vyhledávacích programů na počítači. Soubory vědeckotechnických informací v této formě se běžně nazývají magnetopáskové báze dat [1, 2, 3].

V ČSSR se v současné době zpracovává 14 takovýchto magnetopáskových bází dat a jejich souhrnný přehled s uvedením producenta (tvůrce), provozovatele (organizace zodpovědná za využívání v rámci ČSSR), roku počátku odebírání a počtu záznamů ročního přírůstku je uveden v tabulce 1. Z hlediska informací o biotechnologických je možno mezi nimi stanovit toto pořadí důležitosti: 1. CA SEARCH, 2. World Patent Index, 3. FSTA, 4. PAPERCHEM, 5. COMPENDEX, 6. INSPEC. Referátové služby č. 5 a 6 pokrývají tuto problematiku pouze okrajově. COMPENDEX — fermentační zařízení a přídatná zařízení pro fermentační procesy [filtry, odstředivky, separátory atp.]. INSPEC — má v sekci C podsekcí zabývající se řízením technologických procesů počítači. Z tab. 1 je vidět, že v ČSSR není k dispozici magnetopásková verze Biological Abstracts nazývaná BIOSIS Previews, která by doplňovala poměrně úplnou kolekci těchto služeb pokrývajících základní vědeckotechnické disciplíny.

Tab. 1. Přehled zahraničních magnetopáskových bází dat vědeckotechnických informací zpracovávaných v ČSSR

Název	Zaměření	Producent	Provozovatel	V ČSSR od r.	Roční přírůstek (v tis.)
CA SEARCH	chemie	CAS, USA	VÚTECHP, Praha	1971	480
CIN ¹	chemie	CAS, USA	VÚTECHP, Praha	1977	50
WP1 ²	patenty	Derwent, GB	VÚTECHP, Praha	1970	250
PAPERCHEM	papír, celulóza	IPC, USA	VÚPC, Bratislava	1974	12
FSTA ³	potravinářství	IFIS, NSR	VÚPP, Praha	1978	17
Exc. Medica	lékařské vědy	EM, Holandsko	ÚVLI, Praha	1978	250
COMPENDEX	inženýrství	EI, USA	UTRIN, Praha	1974	100
INSPEC	elektrotechnika ⁴	EE, GB	SITK, Bratislava	1974	160
Pollut. Abs.	životní prostředí	CSA, USA	SVOP, Bratislava	1975	7
Envirotopes	životní prostředí	EIC, USA	SVOP, Bratislava	1975	7
AGRIS	zemědělství	FAO, UNESCO	ÚVTIZ, Praha	1983	100
INIS	atomová energie	MAAE, Rakousko	ÚISJP, Zbraslav	1973	65
TITUS	textil	VTDI, NSR	VÚTESP, Praha	1983	10
PASCAL-Geo.	geologie	BRGM, Francie	CGÚ, Praha	1982	50

1) Chemical Industry Notes, 2) World Patent Index, 3) Food Science and Technology Abstracts, 4) Elektronika, počítače, zpracování informací, automatizace

Jeden z hlavních významů referátových služeb je v tom, že zmírňují nepříznivé důsledky snižování počtu odebíraných titulů primárních časopisů, že umožňují rychlé vyhledávání požadovaných informací rozptýlených ve větším počtu primárních zdrojů, a konečně že s jejich pomocí lze provádět analyticko-prognostické studie zabývající se hledáním hlavních směrů vědeckotechnického rozvoje statistickými metodami.

Nově vytvářené referátové služby

Zde se budeme zabývat pouze oblastí biotechnologií. Jak již bylo řečeno v úvodu, reagovaly některé zavedené referátové služby na vzrůstající publikační činnost v oblasti biotechnologických aplikací jednak rozšířením svého okruhu sledovaných časopisů (kterých vznikla v posledních 5 letech řada) a jednak v některých případech přestavbou skladby dělení na sekce a podsekce, aby tak vznikly samostatné útvary anotující tyto vědní disciplíny. U Chemical Abstracts byla takto přebudována sekce 16: Fermentace a bioprůmyslová chemie. (Kromě toho však v rámci 80 sekcí, na které je obsah CA rozdělen, je zhruba 20 více či méně se dotýkajících problematiky biotechnologií.)

Souběžně s tím začaly vznikat specializované referátové časopisy [pracující na stejném principu anotování zvoleného souboru časopisů a jiných typů publikací] zaměřené na biotechnologie. Některé z nich existují jak v tištěné formě, tak ve formě umožňující přístup v režimu on-line prostřednictvím specializovaných databázových center (viz. ref. 1, 2, 3). Do této kategorie lze zařadit například BIOTECHNOLOGY Abstracts společnosti Derwent Public., Ltd. Jiné mají pouze tištěnou formu (Biotechnology Monthly Update). V tabulce 2 jsou uvedeny některé z těchto referátových služeb spolu s názvem producenta, počtem anotovaných časopisů, počtem citací za rok a cenou pro rok 1984. V tabulce jsou zařazeny pouze nejvýznamnější sekundární informační zdroje, které mají širší tematický záběr.

Tab. 2. Přehled vybraných referátových služeb pro biotechnologie

Název	Producent	Cena (1984) US \$	Počet záz./rok (tis.)	Počet anotovaných časopisů
Biotechnology Abstr.	Derwent, GB	590	12	1 000
Current Biotechnol. Abstracts	Royal Soc. of Chem., GB	400	6	200
Pascal Biotechnol.	CNRS, Francie	260	10	13 000
Telegen Reporter	EIC Int., USA	895	6	7 000
Abstracts in BioCom.	IRL Press, GB	170	10	200

Výhodou těchto specializovaných informačních zdrojů oproti tradičním zavedeným službám je jejich výlučná koncentrace na daný předmět a rozdělení do sekcí odpovídající členění biotechnologického výzkumu a aplikací. Jako příklad je možno uvést BIOTECHNOLOGY Abstracts (tabulka 3). Nevýhodou může být v některých případech určitý nestandardní přístup k anotování, vytváření pouze volně tvořených hesel bez použití závazného slovníku deskriptorů (tezauru), neúplný výběr sledovaných časopisů, specializace pouze na určitá teritoria (USA, Evropa, Japonsko) nebo předmětové oblasti (farmacie, zemědělské aplikace, životní prostředí atp.).

Proto je třeba při posuzování a výběru vhodného referátového informačního zdroje pečlivě prozkoumat všechna tato hlediska, porovnat seznamy citovaných časopisů, prověřit jejich úplnost a zaměření zadáním různých typů rešeršních dotazů a porovnáním získaných výsledků (citacích záznamů) pro jednotlivé systémy.

Způsoby využívání

V zásadě je možno rozlišit dva způsoby automatizace při získávání požadovaných informací z referátových služeb: tzv. selektivní distribuce informací (SDI z angl. Selective Dissemination of Information) a on-line přístup k záznamům v bázi dat. (Blížeší popis obou způsobů viz.

Tab. 3. Tematické dělení referátové služby Biotechnology Abstracts

Počet cit. časopisů: 1 000, Počet základních cit. časopisů: 236 Počet citací za rok: 12 000	
Dělení na sekce:	
A — Mikrobiologie, B — Inženýrství, C — Chemie, D — Léky,	A1 — Genetika, A2 — Fermentace B1 — Biochemické inženýrství C1 — Analýza a struktura D1 — Antibiotika, D2 — Hormony, D3 — Interferon, D4 — Vakcíny, D5 — Ostatní léky
E — Zemědělství,	E1 — Biologické regulační látky, E2 — Fixace dusíku, E3 — Pesticidy, E4 — Kultivace in-vitro, E5 — Ostatní
F — Potraviny,	F1 — Potravinové přísady a jednobuněčné bílkoviny
G — Energie, H — Ostatní chemikálie,	G1 — Paliva H1 — Polymery, H2 — Polynenasycené, H3 — Rozpouštědla, H4 — Chemikálie všeobecně
J — Kultivace buněk, K — Biokatalýza,	J1 — Kultivace živočišných buněk, J2 — Kultivace rostlinných buněk K1 — Izolace a charakterizace, K2 — Aplikace
L — Čištění, M — Likvidace odpadů,	L1 — Downstream zpracování M1 — Likvidace průmyslových odpadů

ref. 1, 2, 3.) V obou případech je výsledkem zpracování rešeršního požadavku, formulovaného vhodnou kombinací deskriptorů (klíčových slov), jmen autorů apod., sada informačních záznamů. Tyto záznamy mají charakter buď signální informace (pouze citace článku v časopise, patentu, disertace, sborníku z konference atp.) v případě bibliografické báze dat anebo se získá úplná informace v případě faktografických bází dat (např. fyzikálně chemická data, charakteristiky mikrobiálních kultur, parametry výrobních zařízení atp.). V prvním případě znamená tedy rešerše pouze první krok, po kterém musí následovat získání citovaného primárního zdroje a jeho prostudování (pozn.: systémy automatizované analýzy odborných textů, které by vedly k jejich věcnému zpracování, nejsou zatím ve stadiu praktické použitelnosti).

Hlavním přínosem těchto automatizovaných systémů vyhledávání požadovaných informací na základě rešeršního požadavku zformulovaného v řeči počítačového programu je možnost zpracování velkého objemu dat v relativně krátkém čase za pomoci vyhledávacích termínů (selekčních prvků) kombinovaných mezi sebou logickými operátory (logický součet, logický součin, negace). Toto jsou prostředky, které nemá k dispozici například ruční rešerše. Je však si třeba uvědomit, že se tímto způsobem získá pouze surový materiál, který je třeba ještě kvalitativně a věcně vyhodnotit. Z tohoto důvodu a vzhledem k tomu, že pro práci s automatizovanými systémy je třeba mít určitou zkušenost a zručnost, vkládá se mezi tyto informační prostředky a konečného uživatele (zadavatele) specialista, který má za úkol provádět formulace rešeršních požadavků a analyzovat získané informační záznamy také po stránce obsahové s ohledem na požadavky a odborné zaměření zadavatele. Požadavkem je, aby tento informační pracovník (v anglické literatuře označovaný jako intermediary) měl základní vzdělání v daném oboru a rovněž v oblasti zpracování vědeckotechnických informací. Faktem je, že v ČSSR se připravuje informačních specialistů tohoto typu soustavně neprovádí. Na jejich výchově by se měly především podílet vysoké školy příslušného odborného zaměření, a to jak formou otevření nových studijních směrů, tak také cestou speciálních postgraduálních kursů. Tato situace by se mohla změnit, pokud se přistoupí k vytváření Středisek analýzy vědeckotechnických informací při větších aglomeracích základního i aplikovaného výzkumu.

Pro prognostické studie mají automatizované systémy zpracování vědeckotechnických informací význam především v tom, že umožňují statisticky vyhodnocovat velké objemy literárních dat v relativně krátkém čase při současné možnosti kombinace několika informačních souborů. Důležité jsou především soubory patentových informací, které jsou hlavními nositeli potenciálně realizovatelných inovačních záměrů. Na ně navazují soubory obchodně ekonomických informací umožňující analyzovat

vzájemné vazby mezi jednotlivými výrobci a jejich chování na prodejních trzích. Je však třeba si uvědomit, že doposud používané analytické metody tohoto typu se opírají o sledování chování velkých výrobních společností, které, zejména v oblasti biotechnologií, nejsou nositeli základních invencí. Proto je nejdůležitějším, avšak také nejobtížnějším krokem počáteční výběr objektů, které je třeba sledovat a analyzovat. Je tedy zapotřebí vytvořit základní aparát umožňující provedení výběru biotechnologických informací z dostupných informačních souborů tak, aby je bylo možno statisticky zpracovat.

Literatura

- [1] BUREŠ, R.: Informační prostředky v chemickém výzkumu a praxi, *Chemické listy*, **74**, 1980, 629—639.
- [2] BUREŠ, R.: Přejít od klasických rešeršních postupů k použití počítačů, *Chemické listy*, **75**, 1981, 1202—1211.
- [3] BUREŠ, R. - KOHNOVÁ, Z.: Zavádění online informačních systémů do rešeršní praxe, *Chemické listy*, **75**, 1981, 93.

Bureš, R.: Analýza možností využití automatizovaných systémů vědeckotechnických informací pro rozvoj biotechnologií. *Kvas. prům.* **32**, 1986, č. 4, s. 88—90.

Článek diskutuje možnosti využití referátových informačních zdrojů pro oblast biotechnologií. Pozornost je věnována především automatizovanému zpracování počítači. Rovněž je stručně diskutována možnost využití pro prognostické studie.

Буреш, Р.: Анализ возможностей использования АС научно-технической информации для развития биотехнологий. *Квас. прум.*, **32**, 1986, № 4, стр. 88—90.

Статья занимается возможностью использования реферативных информационных ресурсов для биотехнологии. Внимание есть прежде всего уточнено на автоматизированную обработку с помощью вычислительных машин. То же есть коротко дискутирована возможность использования для прогностических исследований.

Bureš, R.: Analysis of Possible Utilization of Automatic Systems of Scientific Information for Biotechnology Development. *Kvas. prům.* **32**, 1986, No. 4, pp. 88—90.

The article discusses possibility of exploitation of referal information sources for biotechnology. The primary attention is devoted to computer processing. Also is briefly discussed possibility of their use for prognosis.

Bureš, R.: Analyse der Applikationsmöglichkeiten der automatisierten Systeme wissenschaftlich-technischer Informationen für die Entwicklung der Biotechnologien. *Kvas. prům.* **32**, 1986, Nr. 4, S. 88—90.

Im Artikel wird die Möglichkeit der Ausnutzung der referierten Informationsquellen für das Biotechnologiegebiet diskutiert. Die Aufmerksamkeit wird vor allem der automatischen Verarbeitung durch die Hilfe der Zähler gewidmet. Auch die Möglichkeit der Ausnutzung für die Prognosenzwecke wird kurz angedeutet.