

Perspektivy automatizace dokumentační práce v mikrobiologii a imunologii

Ladislav Hejdánek

in: Časopis český lékař, 1963, p. 582–585

in: Časopis český lékař, 1963, str. 582–585

◆ článek, česky, vznik: srpen 1962?

◆ article, Czech, origin: srpen 1962?

Perspektivy automatizace dokumentační práce v mikrobiologii a imunologii¹ [1962]

Ústav epidemiologie a mikrobiologie v Praze, ředitel prof. dr. K. Raška, DrSc.

V roce 1958 byl na 7. mezinárodním mikrobiologickém kongresu ve Stockholmu vytvořen Stálý výbor pro mikrobiologickou a imunologickou dokumentaci o čtyřech prozatímních členech. Koncem r. 1959 byl vystřídán sedmičlenným výborem, jmenovaným výkonnou radou Mezinárodní asociace mikrobiologických společností (ECIAMS). V průběhu let 1960–1961 byl výbor rozšířen na 10 členů (z ČSSR prof. K. Raška, ze SSSR prof. V. M. Ždanov, jinak jsou zastoupeny: Belgie, Brazílie, Francie, Izrael, Japonsko, Spojené státy americké, Švédsko a Velká Británie), k nimž se druží dalších 20 dopisovatelů, takže koncem r. 1961 bylo ze 34 mikrobiologických společností celých 30 v přímém kontaktu s dokumentačním výborem. Výbor, který se stal mezinárodním členem FID, byl ustaven nikoliv jako pouhý poradní sbor, ale jeho posláním je převzít iniciativu jak v teoretickém studiu, tak v podněcování a organizování praktických pokusů, a ukáže-li se potřeba, v rozvinutí nových systémů a služeb v oboru mikrobiologické a imunologické dokumentace. Zatímco první dva roky byly dobou příprav, získal výbor počátkem r. 1960 trvalé sídlo v UNESCO v Paříži a začal fungovat nejprve prostřednictvím oběžníků (Newsletters). Za předposlední rok však dosáhly přípravy takového stadia, že bylo možno přistoupit k diskusi o programu dlouhodobé činnosti výboru. To se mělo stát námětem první pracovní schůze, která se konala v říjnu 1961 v UNESCO v Paříži.

Mezi materiály z této říjnové schůze nás mohou zajímat zvláště dva elaboráty předsedy Stálého výboru, dr. Pierra R. Brygoo, totiž jednak „Návrh programu“, jednak jeho referát „Použití strojů ve vědecké dokumentaci.“

První z nich byl vypracován na výzvu výkonné rady Mezinárodní asociace mikrobiologických společností a stal se podkladem jejího jednání na londýnské schůzi v září 1961, které vyústilo ve významnou rezoluci 2. října, v níž jsou předseda IAMS, profesor Stuard Mudd, a předseda dokumentačního výboru, dr. Pierre Brygoo, pověřeni ustavením pracovní skupiny, která bude oprávněna jménem IAMS oficiálně zkoumat možnosti a připravovat uskutečnění Mezinárodního střediska pro automatizaci a mechanizaci dokumentace v oborech významných pro mikrobiologii a imunologii.

„Návrh programu“ sestává ze dvou hlavních částí. První, stručnější, se týká předběžného programu práce, který má krátkodobý charakter a je už uskutečňován. Za zvláštní zmínku snad stojí příprava kompletního seznamu periodik, důležitých pro obory mikrobiologie a imunologie, který byl dohotoven a předložen na 8. mezinárodním mikrobiologickém kongresu, který se konal v srpnu 1962 v Montrealu.

Druhá část je zajímavější, protože představuje program střední délky, totiž na dobu 4 let mezi 7. a 9. mikrobiologickým kongresem, tj. 1962–1966.

Především má být provedeno zkoumání, nakolik jsou primární zdroje informací, tj. práce významné pro uvedené obory, zachyceny a analyzovány existujícími dokumentačními službami, tj. sekundárními periodiky. To má být provedeno tak, že bude pořízen kompletní seznam příslušných článků, publikovaných v časopisech důležitých pro mikrobiologii a imunologii v průběhu jednoho celého roku, a že bude zjištěno u každého článku, byl-li zachycen nebo bylo-li o něm referováno v sekundárních

¹ Zpráva o I. pracovní schůzi Stálého výboru pro dokumentaci v oboru mikrobiologie a imunologie v Paříži, 3.–7. října 1961.

periodikách a v kterých. Na základě tohoto materiálu bude stanoveno procento článků takto zachycených alespoň jedním přehledovým nebo referátovým časopisem, čímž bude vyjádřena míra maximální efektivity dnešního systému. Současně bude stanovena pravděpodobnost, že se primární informace objeví dvakrát nebo vícekrát v sekundárních periodikách, čímž bude opět vyjádřena míra ztrát v současném systému. Kromě těchto předních úkolů bude zjišťováno u jednotlivých služeb zpoždění referencí za původními publikacemi, bude provedena zevrubná analýza podle předmětů, jazyků, zemí a časopisů, která poskytne přesný statistický obraz efektivity, duplikací a mezer v jednotlivých složkách dnešního systému, a konečně při této práci budou zachyceny časopisy, které zatím chyběly v seznamu, který byl vloni pořízen. Vzhledem k této významné službě, která bude vykonána také v zájmu vydavatelů primárních periodik, očekává výbor, že se setká s úplným porozuměním jeho zamýšlená žádost, aby vydavatelé poskytli výboru natrvalo a bezplatně po dvou exemplářích svých časopisů. Tak bude mít výbor k dispozici kompletní sbírku zdrojů primárních informací z oboru mikrobiologie a imunologie, která bude v budoucnu pravidelně zpracovávána vybranou specializovanou dokumentační službou v některý typ „paměti“.

Za druhé má být vypracován geografický přehled základních knihoven, v níž jsou příslušné v seznamu uvedené časopisy k dispozici, se současným označením schopnosti té které knihovny pořizovat fotokopie, mikrofilmy nebo jiné typy reprodukcí na požádání. Tak bude kromě jiných výhod poprvé získán přesný obraz o distribuci primárních pramenů k veřejnému použití.

Tímto způsobem budou zjištěny údaje několikerého druhu. Bude objasněna povaha zdrojů primárních informací, geografické rozložení a dostupnost těchto zdrojů v jednotlivých depozitářích, způsob běžného zpracování původních zdrojů a tak i hodnota dnešní soustavy sekundárních informací.

Zvláštní pozornost je v programu věnována rozvoji nových metod mikrobiologické a imunologické dokumentace se zřetelem k použití kybernetických strojů. V této části programu se odrážejí výsledky studijní cesty z června a července 1961 po Spojených státech, jak o ní referoval dr. Brygoo a která se stala základem referátu o „Použití strojů ve vědecké dokumentaci“, o němž podáme zprávu na závěr.

V programu se konstatuje, že dnešní metody vědecké dokumentace jsou v podstatě stále stejné, jako byly na počátku století. Vědecká dokumentace nedržela krok s obrovským rozvojem samotné vědy. Má-li však splnit úkoly, které jsou na ni dnes kladeny, musí se v co nejširším rozsahu automatizovat. Aby mohl být tento dlouhodobý program splněn, je třeba nejprve stanovit hlavní body postupných přípravných opatření, která mají být realizována v uvedeném čtyřletém období 1962–1966.

Program tedy stanoví, že bude ustavena pokusná skupina, složená z dokumentaristů, techniků a mikrobiologů, která především musí přesně definovat žádaný typ výsledků, za druhé musí příslušný materiál zpracovat ve formě signálů pro kybernetický stroj, a konečně sestavit program pro kybernetické zpracování. Dříve než se rozjede celá akce, je třeba vykonat řadu praktických pokusů. Ty však budou v prvním období představovat jen malou část úsilí, které bude celému řešení věnováno, zatímco převážná část bude mít ráz intelektuální a teoretický.

Kódování omezeného materiálu se má již v pokusném období automatizovat, tj. má být prováděno na perforačních psacích strojích, které současně s psaním textu, ostatně stejně nezbytným pro dokumentační práci, převádějí znění na děrné pásky podle zvoleného binárního kódu. Tato část práce by mohla být vykonána v jednom nebo ve dvou existujících dokumentačních střediskách a nevyžádala by si velkého nákladu. Tak by byl zajištěn základní thesaurus informací, který by mohl být dále pokusně zpracováván. Systém by musel být vypracován tak, aby bylo možno vykonat pokusy na

různých elektronkových strojích, takže by bylo nejlépe zajistit spolupráci s některým velkým střediskem, vybaveným rozsáhlým komplexem kybernetických strojů. Tímto způsobem by bylo zajištěno, že nebude nezrale rozhodnuto ve prospěch určitého stroje dříve, než bude proveden potřebný počet pokusných akcí, a že také program bude možno včas korigovat a upravit tak, aby efektivnost výsledku byla optimální. Protože je pravděpodobné, že pozdější použití různých dokumentačních center jako zdrojů bude muset počítat s tím, že nebude jednota ani v kódech, ani v substrátech pro kódování (děrné štítky, různé typy děrných pásek atd.), uvedený komplex musí být dostatečně vybaven stroji na automatický převod jakýchkoli binárních kódů na jediný, kterého bude používáno k zpracování. Základní série zkoušek bude provedena v uvedeném čtyřletém období.

Jen pro zajímavost uvádíme, že rozpočet na toto období, kdy budou prováděny pouze přípravné práce převážně teoretického rázu a kdy se praktické zkoušky budou provádět na strojích existujících středisek, počítá se sumou 50 000 dolarů.

Program byl schválen jak na londýnské schůzi výkonné rady Mezinárodní asociace mikrobiologických společností, tak na první pracovní schůzi dokumentačního výboru.

Referát dr. Brygoo o použití strojů ve vědecké dokumentaci se řadí k prvním souhrnným pracím tohoto druhu a jeho význam přes určitou míru všeobecnosti spočívá ve střizlivém odhadu možností automatizace dokumentační práce ve všech jejích etapách, zatímco většina autorů, kteří se otázkou již zabývali, zůstává převážně jen u technických problémů automatické selekce. Dr. Brygoo objasňuje nejprve skutečný smysl otázky, co všechno je možno v dokumentační práci automatizovat, a varuje před nekritickým očekáváním zázraků. Především je třeba rozlišovat mezi automatizací provádění předepsaných operací a mezi automatizací programování. Některé operace, jako je třídění, vyhledávání a zařazování, lze snadno mechanizovat; jejich postup může být automatizován zavedením programu a jednotlivé dílčí programy mohou být koordinovány v komplexní programy všeobecné. Avšak čím rozmanitější a složitější jsou operace, které mají být automaticky prováděny, tím složitější musí být celkový program (programování). Složitost může narůst tak, že je pro naši představivost již velmi obtížné rozložit stanovené instrukce na elementární postupy, podřízené určitým pravidlům, takže je možno dosáhnout nebo i překročit praktickou mez. Sestavit program je předpokladem každé automatizace, ale o automatizaci samotného programování nemůže být ani řeči. Někdy ovšem je programování jen otázkou času a technické dovednosti. Tam, kde se stanovení operačního programu stává rutinou, musíme ovšem možnost automatizace připustit a využít jí. Naproti tomu v jiných případech představuje stanovení programu výzkumný problém, vyžadující často značnou předběžnou práci. Lze si ovšem představit systémy, schopné značné míry automatického programování v důsledku jistého stupně nezávislosti na operátoru, a takové systémy budou pravděpodobně v budoucnosti realizovány. Nicméně žádný systém není schopen se vymknout svému obecnému programu. Dr. Brygoo proto uzavírá, že automatizace je možno dosáhnout vždy jenom jako výsledku prvotně intelektuální práce, pro niž je stroj pouhým nástrojem, nikoli však iniciátorem. Také v dokumentaci musí být nejprve vykonána velká práce intelektuální jako bezpodmínečný předpoklad veškeré automatizace. A právě po této stránce zůstávají otevřeny obrovské problémy, které nebyly dosud řešeny a které z velké části nebyly dosud ani studovány. Mezi nimi jsou především technické problémy strojního vybavení. Jiné se týkají teoretických otázek vztahu přirozené řeči a řeči stroje. Konečně tu jsou i problémy vztahu mezi vědeckým myšlením určitého oboru a možnostmi aplikace strojů.

Všemi těmito aspekty se dr. Brygoo ve svém referátu postupně zabývá. Pokud jde o technickou stránku strojů, začínají problémy hned při vstupu. Informace, které mají být zpracovány, musí být nejprve převedeny v binární signály.

Náročnost této podmínky je tak velká, že jsou tím kladeny meze rozsahu textů. Kromě toho však vznikají potíže s nutnou redukcí informace tím, že obětujeme mnoho z formy původní (např. tištěné) zprávy; tak např. v binárním přepisu ztrácíme rozdíly mezi malými a velkými písmeny, různými typy použitého písma (třeba verzálka), ztrácíme diakritická znaménka, řecká písmena a jiné speciální znaky. To všechno znamená hned na začátku ztrátu informace. Něco z toho lze odstranit použitím psacích strojů, které současně děrují pásku, ale meze jsou opět dány rozsahem klaviatury. Je pochopitelné, že jakákoli ztráta informace tohoto druhu nemůže být v dalším zpracování už nikdy nahrazena. To je důvodem pro založení paralelního fondu kompletních informací, které nevstupují jako celek do strojního zpracování a které mohou být kdykoliv později více nebo méně automaticky vyhledány a fotograficky nebo jinak reprodukovány. Tak se dosáhne značných ekonomických úspor za tu cenu, že jsou strojně zpracovány pouze některé znaky původní zprávy (např. pořadové číslo zprávy, jeden nebo více autorů, rok vydání, časopis, několik věcných hesel nebo desetinných znaků, ale také třeba plný titul s úplnou citací, souhrn atd.), přičemž se ovšem velmi rychle dosáhne mezí. Automatické čtení originálních textů se současným strojovým převodem do binárního kódu je zatím ve stadiu výzkumu, nehledě na to, že lze zatím uvažovat jen o čtení standardního písma, nikoli používaných typograficky tak rozmanitých druhů písem. Lze však perspektivně předpokládat, že všechny těžkosti, o nichž dosud byla zmínka, budou v budoucnu překonány.

Jinou vážnou otázkou jsou zásoby informací v paměti stroje. Děrné štítky, děrné pásky, magnetické pásky apod. představují paměť neomezené kapacity, ale poměrně pomalého zpracování vzhledem k jeho mechanické povaze. Určitá odpomoc je buď v současném provádění řady operací v průběhu jediného čtení, anebo rozdělení paměti do předtříděných sekcí, aby se zamezilo čtení celé paměti při každém úkonu. Oba postupy mohou být kombinovány, takže různé části paměti mohou být současně čteny z řady aspektů. Naproti tomu nákladnost multidimenzionálních pamětí zatím prakticky vylučuje jejich použití pro dokumentační účely. S jinými druhy pamětí je možno počítat zase jen výhledově. V současné době zůstává nejvhodnějším typem paměti asi magnetická páska (popřípadě bublen).

S velkými problémy se setkáváme rovněž při výstupu. Výsledky zpracování informací jsou automaticky převáděny z binárního kódu v tištěný text. Mechanická povaha této závěrečné operace však opět působí velké omezení rychlosti. I když v rychlostech tištění výsledků strojního zpracování informací bylo dosaženo pozoruhodných úspěchů (600 řádek po 120 typech za minutu je už běžná rychlost, a v některých případech se už dosahuje i rychlosti 1000 řádek za minutu), které ovšem postačují např. pro zhotovení indexů nebo shrnutí výsledků, ale nikoli pro reprodukci celých textů dokumentů. V této věci lze leččeho dosáhnout použitím celé baterie současně tisknoucích strojů. Také bude jistě již v dohlednu možno použít při výstupu sazečského stroje. Byly však vykonány také pokusy s použitím optických a fotografických procesů namísto dosavadních mechanických, takže lze očekávat dosažení rychlostí nedostupných metodám mechanického sázení.

Neméně významná je další řada problémů, jejichž povaha je lingvistická a sémantická. Na této úrovni už nejde jen o převádění písmen (tj. abecedních signálů) na binární signály a naopak, nýbrž o kód, který jednotlivým signálům a skupinám signálů teprve dodává smysl a který z nich dělá nositele informace, tj. řeč (jazyk). Dosavadní jazyky, ať už přirozené nebo umělé, byly vždy lidské; nyní však bylo nutno zahájit studium otázky, jak je možno lidský jazyk přeložit do „jazyka“ interpretovatelného

strojem, pracujícím na binárním základě. Tím teprve může být řešena další praktická otázka překladu jedné lidské řeči do jiné. Zatím však co otázka transformace informací z jedné umělé „řeči“ o známém kódu do jiné takové řeči je v zásadě snadná, stojíme u každé lidské řeči před systémem, jehož kód není exaktně definován. Protože lingvistické obtíže představují jednu z nejvýznamnějších překážek vědecké dokumentace, je automatická dokumentace prakticky neoddělitelná od automatického překládání. Je však neoddělitelná nejen z těchto důvodů, ale jak naznačeno, ze samé podstaty věci i tam, kde zůstáváme u jediného lidského jazyka. Dokud tato otázka nebude vyřešena, nelze očekávat podstatný pokrok nejen v automatickém překládání textů, ale ani v jejich automatické analýze, umožňující automatické zpracování.

Nejvýznamnější skupinou otázek jsou však otázky vztahů mezi vědeckým myšlením a použitím strojů. Velmi mnoho uživatelů informací si zvyklo v selekci (tj. v produkci katalogů, indexů apod.) vidět cíl dokumentační činnosti. Není pak divu, že takové zkreslené představy vedou k názorům, že použití velkých elektronkových strojů je zbytečné a že stejně uspokojivých výsledků je možno dosáhnout nesrovnatelně méně nákladnými prostředky. Jistě je pravda, že jediným kladem použití těchto strojů k selekci informací by byla jistá úspora času, která by však asi nevyvážila vysoké náklady na jejich pořízení a provoz.

Avšak taková redukce dokumentační činnosti není oprávněná. Nejde o pouhou selekci, ale o přímé proniknutí k samotné zprávě, o analýzu jejích dat, o provedení řady operací, které nejen umísťují jednotlivé kódované, analýzou vydělené informace do paměti, ale které jsou schopny je logicky zpracovat v souvislostech určených rámcovým programem při dosažení významného stupně sebeprogramování, atd. A k tomu všemu je použití strojů o vysoké kapacitě bezpodmínečně nutné.

Ovšem už získání času je zvláště ve vědě neobyčejně významné. Vědecké informace velmi brzo ztrácejí svou cenu a zůstávají předmětem historie. Ale také opačně záleží na mnohých okolnostech, kdy vědecká informace dosáhne maxima své hodnoty, tj. kdy a v jakém rozsahu začne být prakticky uplatňována. V samotné publikaci má informace jen potenciální hodnotu. Ovšem již při sestavování výsledků experimentální práce (bez ohledu na zpracování literatury) je velmi často třeba sledovat souvislosti fakt neobyčejně pracným způsobem, vyžadujícím nesčetných operací (např. statistických), kdy tempo lidské práce nepostačuje požadavkům výzkumu. Situace v biologii se v mnoha ohledech počíná v této věci přibližovat situaci v matematice a fyzice. Nehledíc už na to, že na pravděpodobnostním základě bude nutno vybudovat také metody, umožňující hodnotit vědecká data, neboť informaci musíme podle Shannona matematicky formulovat jako nepravděpodobnost, resp. jako negativní entropii.² Kdyby stroje na základě statistické analýzy mohly rozlišovat mezi cennými nebo málo cennými až bezcennými informacemi, mělo by to význam nejen pro vhodné poskytování již existujících literárních informací, ale také pro vymezení významných dosud neprobádaných oblastí a tak pro zaměření vědecké práce.

Avšak nejde jen o přístup z hlediska pravděpodobnosti, ale také o logické zpracování, kterého jsou stroje rovněž schopny. Stroje tak mohou provádět určité úsudky a formulovat na základě množství dat jakési schematické hypotézy. Takové operace jsou jistě možné v jednoduchých uzavřených systémech, ale zůstávají zatím otázkou u otevřeného systému, jakým je lidské poznání. Avšak i když snad takové myšlenky mohou v tuto chvíli vypadat jako sny a fantazie, jsou to sny, o kterých lze vážně

² O entropii se dnes v informační teorii mluví jako o míře neurčitosti (nějaké zprávy), znamená však obecně také míru pravděpodobnosti daného stavu nějaké soustavy (ať už hypotetické nebo reálné, např. fyzikální, biologické apod.).

diskutovat. Tyto možnosti se mohou jako schůdné nebo neschůdné ukázat jen v praktickém použití strojů, a jejich použití v dokumentační činnosti se může stát významným stimulem k dalším pokusům. Otázkou není jen nahradit lidi a dělat totéž, jenomže ve větším a širším měřítku. Jde o to, dělat cosi jiného, něco víc, dosáhnout kvalitativního pokroku.

Zůstaneme-li však u praktických hledisek, musíme přiznat, že už částečná mechanizace, i když by se zdála velmi skromnou a nepříliš efektivní, může představovat nezbytnou průpravu k další práci, jejímž cílem je úplná mechanizace. Nelze však v době obrovského urychlování technického pokroku čekat na pomalé přirozené dozrávání nových metod. Nejde o ustavičné pokusy dosáhnout nové rovnováhy, protože jsou vždy znovu kompromitovány novými objevy. Naší povinností je předvídat budoucí směry vývoje a šetřit tak čas a peníze z hlediska dlouhodobých perspektiv. Ovšem shledávání všech těchto hledisek působí, že jakýkoli projekt je velice obtížný a vyžaduje důkladné přípravy výzkumného charakteru. Tyto přípravy nemohou vykonat jen samotní technické, ale ve spolupráci s nimi i specialisté příslušných oborů vědecké dokumentace (lingvistiky, logiky, matematiky a tzv. informační inženýři) — a samozřejmě i vědečtí odborníci těch oborů, které mají být informačně zpracovány. Tak se ukazuje nutnost vytvořit celou výzkumnou skupinu.

To všechno přesahuje obvyklé hranice vztahů mezi tím, kdo zpracovává informace, a mezi jejich konzumenty. Tady nejde o komerční podnik, ale o vědeckou výzkumnou práci, která je sice zaměřena k praktickému cíli, ale momentálně je osvobozena od hledisek obvyklých při výrobních procesech. Bez vytvoření takových výzkumných skupin, bez založení experimentálních středisek není možno očekávat pokrok v stále potřebnějším zvyšování podílu dokumentační práce na vědeckém výzkumu i na uplatňování jeho výsledků v praxi.

Do redakce došlo v srpnu 1962.