



Ministerstwo Nauki
i Szkolnictwa Wyższego



Doskonała
Nauka

IV Konferencja Naukowa Konsorcjum BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”

RAPORT I REKOMENDACJE

Elżbieta Nazaruk

Karolina Famulska-Ciesielska

Politechnika Lubelska
2025-03-28

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Doskonała Nauka II – Wsparcie konferencji naukowych



Spis treści

1. Wprowadzenie	2
2. Przebieg konferencji	3
I. Dzień pierwszy	3
II. Dzień drugi	10
3. Wnioski	18
4. Rekomendacje	20



1. Wprowadzenie

W dniach 26–27 września w Politechnice Lubelskiej zorganizowano czwartą edycję konferencji konsorcjum BazTech. Organizatorem wydarzenia było Centrum Informacji Naukowej i Technicznej (CINT) PL, a wsparcie finansowe uzyskano z Ministerstwa Nauki i Szkolnictwa Wyższego w ramach programu „Dokształcająca nauka II – Wsparcie konferencji naukowych”. Celem konferencji była dyskusja na temat obecnego stanu polskich bibliograficznych baz danych oraz scenariuszy ich dalszego rozwoju. Integracja środowiska twórców baz danych i wspólne wyznaczanie kierunków dalszego rozwoju powinny umożliwić – zgodnie z założeniami organizatorów – zwiększenie powszechnego i otwartego dostępu do publikacji w polskich czasopismach naukowych. Konferencja została skierowana do twórców, dystrybutorów i użytkowników dziedzinowych serwisów bibliograficznych oraz badaczy zajmujących się szeroko pojętą komunikacją naukową, a jej tematyka miała uaktualnić i poszerzyć katalog zagadnień poruszanych na poprzednich [konferencjach Konsorcjum BazTech](#).

Zapis wideo z konferencji dostępny jest na kanale Politechniki Lubelskiej na YouTube:

- Dzień 1: <https://www.youtube.com/live/2ICCFn5jdb4>,
- Dzień 2: <https://www.youtube.com/live/im3jrQFS3NM>.

W spotkaniu poświęconym bibliograficznym bazom danych i ich funkcjonalności w dzisiejszym świecie uczestniczyło 130 przedstawicieli 45 polskich instytucji naukowych. Wśród uczestników znaleźli się m.in. bibliotekarze, wydawcy, muzealnicy, pracownicy instytutów naukowych i naukowcy reprezentujący różne dyscypliny.

Wydarzenie objęte było patronatem honorowym JM Rektora Politechniki Lubelskiej oraz Konferencji Dyrektorów Bibliotek Akademickich Szkół Polskich. Patronat medialny zapewniły Stowarzyszenie EBIB, Forum Akademickie oraz Pollub.tv.

Podczas konferencji uczestnicy mogli oglądać postery prezentujące bazy danych przedstawiane w poszczególnych sesjach. Znalazły się na nich podstawowe informacje o bazach, jak również treści, które stanowiły uzupełnienie wystąpień.

Wsparcia finansowego konferencji udzielili partnerzy merytoryczni, prezentujący swoje oferty na stoiskach, a niektórzy w trakcie wystąpień. Były to następujące firmy: ARFIDO, EBSCO, Clarivate, HAN, Wydawnictwo Naukowe PWN, Aleph Polska, Digital Center i Wydawnictwo Naukowe Helion.



2. Przebieg konferencji

I. Dzień pierwszy

Otwarcia konferencji dokonali Prorektor ds. Nauki Politechniki Lubelskiej prof. dr hab. inż. Wojciech Franus wraz z prof. Uniwersytetu Zielonogórskiego podsekretarz stanu w MNiSW dr hab. inż. Marią Mrówczyńską oraz dr Katarzyną Weinper – Dyrektor CINT. Sesję pierwszą „Bibliograficzne bazy danych – krajobraz, tendencje i wyzwania”, którą poprowadził dr Henryk Hollender (Stowarzyszenie EBIB), zainaugurowało nagranie wystąpienia pt. *Bibliograficzne bazy danych kontra wyszukiwarki internetowe – analiza jakości informacji, funkcjonalności i użyteczności* dr hab. Małgorzaty Kowalskiej-Chrzanowskiej, prof. UMK (Uniwersytet Mikołaja Kopernika w Toruniu). Prelegentka skupiła się na porównaniu obydwu narzędzi wymienionych w tytule i oceniła ich przydatność w gromadzeniu literatury naukowej, biorąc pod uwagę m.in. interfejs użytkownika, głębokość indeksowania oraz dostępność i wiarygodność oferowanych informacji. Przeanalizowane zostały bazy BazTech, BazEkon i BazHum oraz wyszukiwarki Google, Google Scholar, Yahoo i Bing. Jako największe zalety wyszukiwarek prelegentka wymieniła rozpiętość tematyczną i szeroki zakres treści, a w przypadku baz danych – poziom merytoryczny indeksowanych treści czy dostępność bibliografii załącznikowych i streszczeń indeksowanych artykułów. Autorka stwierdziła, że wyszukiwarki wciąż nie są w stanie w pełni zastąpić bibliograficznych baz danych, choć na niekorzyść tych drugich działają np. konieczność logowania w niektórych przypadkach, a także wciąż obecne (choć rzadziej niż w wyszukiwarkach) błędy i niespójności bibliograficzne.

Na zakończenie podkreśliła znaczenie baz danych w ewaluacji czasopism i pracowników naukowych, jak również ich przydatność do analizy wskaźników naukometrycznych.

Następnie prof. dr hab. Marek Niezgódka (Centrum Technologii i Nauk Obliczeniowych Politechniki Śląskiej) mówił w wystąpieniu o Infrastrukturze komunikacji naukowej w epoce sztucznej inteligencji, rozpoczynając od stwierdzenia, że otwarta nauka to oczywistość i powinność, a „zamykanie” jej, czyli udostępnianie prac naukowych wybranym użytkownikom (np. za opłatą), jest zjawiskiem wtórnym. Prelegent opisał kilka inicjatyw promujących w Europie diamentowy model OA. Podkreślił rolę finansowanego przez UE projektu DIAMAS, którego celem jest zwiększanie widoczności otwartych zasobów oraz określanie standardów otwartej nauki. Zwrócił jednocześnie uwagę na pojawiające się systemy, np. DORA (San Francisco), EUA (EU) czy COARA, które monitorują funkcjonowanie wyszukiwarek i oceniają jakość informacji przekazywanych przez sztuczną inteligencję (AI). Prof. Niezgódka zaznaczył, że na gruncie polskim konieczna jest harmonizacja i stawianie na wielodyscyplinowość oraz tworzenie platform integrujących bazy typu BazTech, BazEkon, BazHum.



Zwrócił także uwagę na istotną rolę bibliotek i innych instytucji w dokumentowaniu i udostępnianiu informacji oraz w zarządzaniu metadanymi. Podsumowując swoje wystąpienie, zauważył, że aby osiągnąć cel, którym jest dotarcie do nowoczesnego, przejrzystego modelu nauki, należy zintegrować AI z infrastrukturą badawczą oraz otwartymi zasobami.

Druga sesja „Gromadzenie i opracowanie danych do baz i serwisów agregujących – przegląd praktyk”, moderowana przez Dorotę Buzdygan z Politechniki Krakowskiej, rozpoczęła się wystąpieniem Doroty Wojewody z Biblioteki Głównej Uniwersytetu Ekonomicznego w Poznaniu. Prelegentka przedstawiła stan obecny i perspektywy rozwoju bibliograficzno-pełnotekstowej bazy BazEkon, rejestrującej zawartość wydawanych w Polsce czasopism i wydawnictw z zakresu nauk ekonomicznych i pokrewnych. Baza, która powstała z inicjatywy Biblioteki Głównej Uniwersytetu Ekonomicznego w Krakowie w 2011 r. skupia dziś 7 bibliotek uniwersyteckich należących do konsorcjum. Obecnie współpracuje z bazą 35 redaktorów, a liczba indeksowanych czasopism to 697. BazEkon wyróżnia się spośród innych baz tym, że indeksuje nie tylko czasopisma naukowe, ale i fachowe (związane z różnymi branżami). Baza wciąż się rozwija, w 2024 r. przybyło 10 nowych tytułów. Jeśli chodzi o plany i wyzwania na przyszłość, prelegentka zwróciła uwagę na potrzebę zewnętrznego wsparcia BazEkon w procesach aktualizacji oprogramowania i opisów bibliograficznych, a także archiwizacji i udostępniania danych.

Następnie Karolina Tabak z Muzeum Historii Polski zaprezentowała Bazę Czasopism Humanistycznych i Społecznych BazHum. Prelegentka poinformowała, że aktualnie BazHum jest w stanie zamrożenia – od 2020 r. nie są rejestrowane nowe publikacje. Scharakteryzowała dwa serwisy BazHum – bazę bibliograficzną oraz pełnotekstową. Pierwszy ze wspomnianych serwisów jest obszerniejszy i obejmuje 900 tytułów, a drugi – bardziej szczegółowy i uwzględnia 400 pełnotekstowych czasopism. W BazHum dominują teksty z dziedziny historii, które stanowią aż 40%. Zdaniem prelegentki największymi zaletami BazHum są otwartość zasobów i wolna dystrybucja metadanych. Został też przedstawiony plan rozwoju bazy. Mieści się w nim ponowne jej uruchomienie (odmrożenie) oraz scalenie dwóch funkcjonujących odrębnie serwisów. Miałoby to usprawnić proces publikacji i ułatwić wyszukiwanie rekordów przez uproszczenie interfejsu strony. Jako wyzwanie przedstawiona została także konieczność dostosowania bazy do potrzeb osób z niepełnosprawnościami oraz nawiązanie współpracy z innymi bazami czy repozytoriami.

Kolejną bazę danych – BazTech, z zakresu nauk technicznych, ścisłych i ochrony środowiska – zaprezentowała Stanisława Pietrzyk-Leonowicz (CINT PL), która mówiła o partnerstwie 22 instytucji,



głównie polskich uczelni technicznych, działających od 25 lat w Konsorcjum

BazTech. Prelegentka zaprezentowała zasady działania konsorcjum, w tym jego strukturę, procedury komunikacji, zakres umowy o współpracy. Zwróciła uwagę na zmianę, jaka zaszła w modelu realizacji projektu dofinansowanego przez MNiSW. Politechnika Lubelska, która jest koordynatorem BazTech od 2022 r., przyjęła model lider-partner, co zacieśniło współpracę zarówno między pełnomocnikami, jak i uczelniami. Mówiąc o perspektywie rozwoju BazTech, referentka podkreśliła m.in. wagę starań o zapewnienie użytkownikom i twórcom bazy nowych rozwiązań technologicznych, szerszej dostępności i lepszej widoczności.

Po wystąpieniu przedstawicielki Politechniki Lubelskiej głos zabrała Małgorzata Grabowska z Biblioteki Geologicznej Państwowego Instytutu Geologicznego w Warszawie. Prelegentka zarysowała obecny stan Bibliografii Geologicznej Polski (BGP), przedstawiającej dorobek naukowy środowiska polskich geologów oraz przedstawiła plany na przyszłość. BGP ukazuje się od 1922 r. w wersji papierowej, a od 1989 r. tworzona była w oparciu o program CDS/ISIS. W 2006 r. rozpoczęto opracowywanie BGP w systemie ALEPH. Referentka poinformowała, że w 2020 r. przeprowadzono konwersję rekordów z wcześniej używanego ISIS do ALEPH, czyli zintegrowanego komputerowego systemu bibliotecznego, co znacznie usprawniło pracę zespołu. Za jedno z największych wyzwań BGP prelegentka uznała konieczność dzielenia wielu zadań w niewielkim zespole biblioteki.

Ostatnią prelegentką pierwszej części w sesji drugiej była dr Aneta Drabek z Biblioteki Uniwersytetu Śląskiego. To wystąpienie poświęcone było analizie polskich czasopism indeksowanych w CEJSH (The Central European Journal of Social Sciences and Humanities), czyli bibliograficzno-pełnotekstowej bazie z zakresu nauk humanistyczno-społecznych. Pani Drabek zwróciła uwagę na fakt, że w bazie odnaleźć można przede wszystkim teksty w języku polskim, mimo że nazwa wskazuje na obszar obejmujący Europę Środkowo-Wschodnią. Prezentując obecny stan bazy, prelegentka podkreśliła, że indeksowane w niej 1843 tytuły polskich czasopism są punktowane w wykazie Ministerstwa Nauki i Szkolnictwa Wyższego. Liczba ta stanowi 65% wszystkich zarejestrowanych w CEJSH pism. Prelegentka zaznaczyła, że celem, jaki stawiają przed sobą twórcy bazy jest uzyskanie wyniku 100% punktowanych czasopism. Poinformowała również, że 19% indeksowanych tekstów znajduje się jednocześnie w Scopusie, a 34% w bazie DOAJ. Nawiązując do zawartości tematycznej bazy, zwróciła uwagę, że najwięcej, bo aż 81% tekstów to prace z zakresu prawa kanonicznego, a wśród czasopism dominują tytuły z dziedziny polonistyki, historii i nauki o kulturze i religii. Pod koniec wystąpienia skupiła się na wymienieniu zalet i wad omawianej bazy. CEJSH wyróżnia spośród innych baz to, że jest to jedyna polska

Strona 5 z 20

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Doskonała Nauka II – Wsparcie konferencji naukowych



baza z zakresu nauk humanistyczno-społecznych. Mówiąc zaś o problemach związanych

z funkcjonowaniem CEISH, prelegentka wspomniała o pojawiających się opóźnieniach w publikacjach, przestarzałym oprogramowaniu oraz obecności zduplikowanych rekordów.

Drugą część sesji poprowadziła Lidia Derfert-Wolf – koordynatorka merytoryczna bazy Baz-Tech, która zaprosiła do wygłoszenia referatu Dorotę Ubysz z Główniej Biblioteki Lekarskiej. Referentka mówiła o funkcjonowaniu, możliwościach oraz oczekiwaniach związanych z Polską Bibliografią Lekarską (PBL), która działa od 1979 r. Przedstawiła metody wprowadzania, przetwarzania, wyszukiwania oraz udostępniania danych. Zaprezentowała również sposoby wykorzystania bazy przez użytkowników Główniej Biblioteki Lekarskiej, m.in. do tworzenia tematycznych zestawień literatury fachowej. Podczas wystąpienia pani Derfert-Wolf nawiązała do języka informacyjno-wyszukiwawczego Medical Subject Headings (MeSH), na podstawie którego na gruncie polskim została opracowana baza Tez-MeSH, zawierająca dodatkowo terminologię polską z zakresu medycyny, nauk pokrewnych i organizacji ochrony zdrowia. Aktualna wersja bazy zawiera około 30 000 rekordów. Prelegentka wskazała cele przyświecające osobom skupionym wokół PBL: ułatwienie dostępu do bazy w zakresie przeszukiwania danych, zwiększenie możliwości udostępniania zasobów oraz poszerzenie jej rekordów o źródła historyczne oraz nowe indeksacje.

W kolejnym wystąpieniu dr Beata Koper i dr Cezary Rosiński z Instytutu Badań Literackich PAN (IBL) przedstawili sytuację i kierunek rozwoju Polskiej Bibliografii Literackiej (PBL). Prelegenci mówili o konieczności i zasadności automatyzacji prac dokumentacyjnych PAN-u, czyli pozyskiwania, przetwarzania i publikacji danych bibliograficznych. Automatyzacja polega m.in. na selekcji danych z baz Biblioteki Narodowej, konwersji formatu MARC21 do struktury danych PBL i prezentacji danych w używanych powszechnie formatach bibliograficznych (Bibtex, RIS). Podczas wystąpienia omówione zostały też dotychczasowe rozwiązania technologiczne przyspieszające prace nad bazą danych oraz metody pozyskiwania danych wykorzystujące sztuczną inteligencję. Przedstawiciele IBL-u przypomnieli słuchaczom o tym, że PBL to nie tylko zbiory strictly literackie. Do zasobów bazy wchodzi, oprócz tekstów z literatury polskiej, tłumaczenia, nagrania audycji, rekordy spektakli i filmów. Prelegenci zwrócili uwagę na nadrzędny cel automatyzacji bazy PBL. Jest nim potrzeba aktualizacji i szybkiego upowszechniania ogromnych zasobów literackiego rynku wydawniczego. PBL jest źródłem szeroko wykorzystywanym przez naukowców, a powiązanie jej z API BN oraz z możliwościami AI pozwala usprawnić pracę i sprawnie wyszukać pożądane informacje.

Strona 6 z 20

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Doskonała Nauka II – Wsparcie konferencji naukowych



Następnie zaprezentowana została unikalna w skali europejskiej baza Materia Medica Opoliensis (MMO), gromadząca informacje z zakresu wiedzy o naturalnych substancjach leczniczych. Dane te pochodzą z czasopism poświęconych ziołarstwu i ziołolecznictwu. Projekt jest realizowany na Uniwersytecie Opolskim w Centrum Nowoczesnych Technologii i w Bibliotece Uniwersytetu Opolskiego, a zainicjował go Zbigniew Skotnicki. Prelegentki, Danuta Szewczyk-Kłos i Dorota Wierzbicka-Próchniak, przedstawiły cel projektu, jakim jest stworzenie narzędzia badawczego do wykonywania m.in. analiz porównawczych naturalnych substancji. Zaprezentowane zostały możliwości użycia bazy w procesach dydaktycznych i naukowych, jak również plany na przyszłość związane z rozwojem MMO. Głównym celem działania bazy jest wsparcie pracowników medycznych i farmaceutów, polegające na analizie działań substancji, projektowaniu kompozycji leków naturalnych czy też ułatwianiu znalezienia odpowiednich lekarstw ziołowych. Docelowymi odbiorcami bazy są medyczne ośrodki akademickie. Mimo że MMO cały czas się rozrasta, nie jest szeroko udostępniana ze względu na bezpieczeństwo i możliwości nadużyć w wyniku wykorzystania dostępnych tam informacji przez nieodpowiedzialne podmioty.

W dalszej kolejności wystąpiła Monika Małecka-Krawczyk reprezentująca Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego (ICM). Tematem referatu była otwartość Biblioteki Nauki działającej w ramach Platformy Otwartej Nauki ICM UW. Jako największa polska biblioteka cyfrowa ma ona na celu otwarte udostępnianie pełnotekstowych publikacji naukowych (artykułów z czasopism, książek) ze wszystkich dziedzin i dyscyplin. Prelegentka zwróciła uwagę, że gromadzenie i opracowywanie danych w Bibliotece Nauki zależne jest od warunków związanych z dywersyfikacją źródeł metadanych, uwarunkowań prawnych i różnicowania modeli współpracy z partnerami. Jako cel działań serwisu wskazała promocję otwartej nauki, a także dążenie do transparentności i jednoznaczności w określeniu praw do danego tekstu (otwarte licencje).

Na zakończenie tej sesji ponownie wystąpiła dr Drabek, która tym razem mówiła o bazie czasopism elektronicznych Arianta, zadając pytanie o zasadność jej funkcjonowania w dzisiejszym świecie. Baza powstała 20 lat temu w związku z potrzebą zgromadzenia w jednym miejscu informacji o polskich naukowych czasopismach elektronicznych. Nielicznie reprezentowane wówczas e-czasopisma miały swoje miejsce na stronach instytucji sprawczych i często niełatwo było do nich dotrzeć. Celem Arianty od początku była aktualizacja i uzupełnianie profili pism. Do dziś pozostaje ona największą bazą polskich bieżących czasopism naukowych w kraju. Korzystają z niej naukowcy i studenci, traktujący Ariantę jako



sprawdzone źródło danych, umożliwiające szybkie dotarcie do tekstów. W podsumowaniu prelegentka stwierdziła, że baza wciąż pozostaje potrzebnym i wiarygodnym źródłem informacji o zawartości czasopism, licencjach, na których udostępniane są konkretne tytuły, jak również o wskaźniku impact factor skatalogowanych artykułów.

Po zakończeniu sesji rozpoczął się panel dyskusyjny zatytułowany „Współpraca w relacjach: wydawca – redakcja – dziedzinowe bazy danych – serwisy agregujące”, który moderowała Stanisława Pietrzyk-Leonowicz koordynatorka Konsorcjum BazTech. Do rozmowy zostali zaproszeni: dr Michał Doligalski z Departamentu Innowacji i Rozwoju MNIŚW, dr Aneta Drabek reprezentująca tu bazę CEJSH, prof. Arkadiusz Gola – redaktor czasopisma „Applied Computer Science” (ACS) z Politechniki Lubelskiej, Beata Klyta z Uniwersytetu Śląskiego, reprezentująca Stowarzyszenie Wydawców Szkół Wyższych, Dorota Wojewoda z bazy BazEkon oraz Dorota Bazan z Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego.

Na wstępie twórcy baz danych zostali poproszeni o sformułowanie oczekiwań wobec wydawców i redakcji czasopism. Aneta Drabek wskazała na potrzebę opracowania pełnych i rzetelnych metadanych, dzięki którym możliwa będzie identyfikacja każdego artykułu. Panelistka mówiła także o istotności dostępu do pełnych tekstów w sieci oraz o konieczności bieżącego indeksowania nowo wydanych numerów. Odniosła się także do dynamicznej sytuacji w rozwoju czasopism i stwierdziła, że praktyki sprzed 15 lat często są już nieaktualne. Zgadzając się z wypowiedzią przedmówcy, głos zabrala pani Wojewoda, podkreślając potrzebę dbałości o jakość danych i współpracę z wydawcami. Zwróciła też uwagę na zasadność archiwizowania artykułów w sieci i dbania o to, by nie zginęły bezpowrotnie wraz z dezaktywacją prowadzącego do nich linku. Następnie głos zabrala Dorota Bazan, która wyraziła przekonanie, że warto zachęcać wydawnictwa do publikowania w formule Open Access, na w pełni otwartej licencji. Podkreśliła przy okazji, że dziś częste są sytuacje, gdy pozornie mamy do czynienia z tekstem otwartym, a przy próbie dotarcia do pełnego artykułu okazuje się, że plik tekstowy jest zablokowany lub niekompletny. Postulowała też dbałość o aktualizację metadanych wprowadzanych do baz. Do tego zagadnienia odniosła się B. Klyta, stwierdzając, że powinny istnieć jednoznaczne kryteria, które umożliwią redaktorom właściwe przygotowanie metadanych i uniknięcie błędów. Panelistka wyraziła pewność, że większą wartość ma pełne i poprawne przygotowanie danych na bieżąco niż praca nad poprawianiem tego, co zostało nieodpowiednio wprowadzone. Następnie głos zabrał prof. Gola, podkreślając, jak ważna jest indeksacja czasopism w różnych bazach. Dzieląc się własnymi doświadczeniami jako redaktor pisma ACS zauważył, że indeksowanie czasopism w BazTech, BazEkon czy też innych bazach, ma dużo mniejsze znaczenie i wartość rynkową niż możliwość



wyszukania danej pozycji w Scopus czy Web of Science. Według prof. Goli

dopiero

indeksowanie

w jednej z dwóch wyżej wymienionych baz przynosi czasopismu prestiż, a dostanie się do nich i utrzymanie na ich liście nie jest łatwe i opiera się na konieczności spełnienia szeregu kryteriów. Z kolei dr Doligalski zwrócił uwagę na dwie kwestie. Po pierwsze na konieczność dialogu między twórcami baz bibliograficznych a redaktorami w celu ustalenia kryteriów dobrych publikacji, po drugie na tzw. złoty rekord, który jako wzorcowy zestaw najbardziej aktualnych i wiarygodnych danych dotyczących publikacji, umożliwi wyeliminowanie duplikatów i słabej jakości tekstów. Komentarz dodała także dr Drabek, mówiąc o konieczności stałej współpracy między różnymi bazami oraz Biblioteką Narodową. Panelistka podkreśliła, że słuchacze powinni mieć świadomość, że większość użytkowników, szukając informacji w internecie, korzysta z Google Scholar, a nie z baz specjalistycznych. Jednocześnie bazy te udostępniają dane Google Scholar, z czego wynika, że za pośrednictwem najbardziej znanej wyszukiwarki ludzie korzystają z baz danych w otwartym dostępie. Dr Drabek przestrzegła też słuchaczy przed korzystaniem z drapieżnych baz danych czy stron takich wydawnictw, gdyż jakość publikowanych tam tekstów jest niedostatecznie sprawdzana lub zupełnie nieweryfikowana. Odnosząc się do wypowiedzi przedmówczyni, pani Bazan dodała, że ważne jest indeksowanie w Google Scholar, bo jest to baza szeroko znana oraz, co istotne, darmowa i otwarta, w przeciwieństwie do prestiżowych baz, takich jak Scopus czy WoS.

Następnie prowadząca panel Stanisława Pietrzyk-Leonowicz zadała pytanie o to, dlaczego indeksowanie w bazach jest tak ważne i kto decyduje o aplikacji danego czasopisma do konkretnej bazy. W pierwszej kolejności głos zabrała Beata Kłyta, wskazując, że o aplikacji do danej bazy decyduje redaktor naczelny czasopisma i przy wyborze często bierze pod uwagę czynnik ekonomiczny, decydując się na najtańszą opcję. W odniesieniu do tej wypowiedzi stanowisko zajął też prof. Gola, zwracając uwagę na to, że rozwijające się czasopisma i redakcje na początku drogi aplikują nawet do niezbyt wartościowych, mało sprawdzonych baz. Według redaktora pisma ACS zdecydowanie lepiej być indeksowanym w stosunkowo nielicznych, lecz sprawdzonych bazach. Następnie podjęta została kwestia jakości indeksowanych czasopism i publikowanych w nich artykułów. Do dyskusji zostali zaproszeni także uczestnicy z sali. Dr Henryk Hollender ze Stowarzyszenia EBIB wyraził przekonanie, że w Polsce powinny się pojawić ogólnie obowiązujące, konkretne normy zapisu bibliograficznego, tak jak ma to miejsce na Zachodzie. Dyskutant odwołał się tu do pojęcia „złotego rekordu” przywołanego wcześniej przez M. Doligalskiego. Opracowanie jednolitych zasad umożliwiłoby wydawnictwom utrzymanie publikowanych tekstów na jednakowo wysokim poziomie. Hollender dodał też, że do weryfikacji treści warto wykorzystywać dziś AI. Ważne jednak, by zdawać sobie sprawę, że jakość tej

Strona 9 z 20

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Dobra Nauka II – Wsparcie konferencji naukowych



weryfikacji będzie zależała od precyzji zadanego pytania (promptu). W dyskusji padło też pytanie, jaki jest cel istnienia agregatorów treści. Głos w tej sprawie zabrał prof. Dariusz Czerwiński z Politechniki Lubelskiej i stwierdził, że ich zadaniem jest streszczenie użytkownikowi (naukowcowi, studentowi) treści, których nie chce czy nie potrzebuje czytać w całości. W tym miejscu włączyła się do rozmowy podsekretarz stanu prof. Mrówczyńska, która poinformowała uczestników konferencji, że w ministerstwie trwają prace nad dokumentem regulującym działalność wydawnictw naukowych w związku z polityką otwartych danych. Ważną nowością będzie to, że zdecydowanie większy nacisk ma zostać położony na darmowe, otwarte udostępnianie tekstów naukowych. Na koniec prof. Mrówczyńska dodała, że warto mieć świadomość, iż bibliograficzne bazy danych także wpisują się w otwartą naukę i stanowią tzw. big data do analizy. Po tych słowach głos zabrała Lidia Derfert-Wolf, która zadała panelistom pytanie, czy warto dodawać do artykułów naukowych link kierujący do danych badawczych. Jednocześnie stwierdziła, że według niej jest to pozytywne zjawisko, gdyż pozwala na dotarcie do tekstów źródłowych. Do wypowiedzi L. Derfert-Wolf odniósł się prof. Gola i, zgadzając się z tą opinią, przypomniał, że coraz więcej czasopism udostępnia źródła i że NCN (Narodowe Centrum Nauki) także działa zgodnie z polityką otwartości, promując licencję CC BY. Prof. Gola dodał, że bardzo istotną funkcję spełniają dziś identyfikatory DOI artykułów naukowych – łatwe do wygenerowania, a bardzo przydatne w procesie wyszukiwania danego tytułu. Dyskusja panelowa zakończyła pierwszy dzień konferencji.

II. Dzień drugi

W drugim dniu odbyły się trzy sesje. Pierwsza z nich, zatytułowana „Uwarunkowania techniczne w rozwoju bibliograficznych baz danych” i moderowana przez dr. Cezarego Rosińskiego z Instytutu Badań Literackich PAN rozpoczęła się wystąpieniem dr. Stanisława Skórki reprezentującego Bibliotekę Główną AGH w Krakowie. Prelegent dokonał analizy porównawczej bibliograficznych baz danych z tzw. heurystykami architektury informacji. Wychodząc od wyjaśnienia nieoczywistych pojęć, dr Skórka poinformował, że architektura informacji (InfoArch) ułatwia organizację danych przez ułożenie opisów w logicznie powiązane kategorie, pozwalające na zidentyfikowanie zasobów oraz szybsze ich wyszukiwanie. Wyjaśnił, że na InfoArch składają się również interfejsy użytkownika oraz systemy nawigacji, ułatwiające przeszukiwanie i kontekstowe łączenie danych. Składnikami, którymi zarządza InfoArch, są: układ metadanych, ich struktura, indeksacja i różne sposoby wyszukiwania (np. temat, słowa kluczowe). Dr Skórka wytłumaczył też znaczenie samego pojęcia heurystyk, które tu rozumiane jest jako zbiór zasad porządkujących tworzenie przejrzystych interfejsów. Dobrze zbudowany interfejs

Strona 10 z 20

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Dobra Nauka II – Wsparcie konferencji naukowych



cechują: spójność, elastyczność, czytelność i dostępność dla użytkownika.

Istotnym wyznacznikiem brany pod uwagę w InfoArchu jest użyteczność wyszukiwanych treści. Prelegent zauważył też, że architektura informacji powinna uwzględniać „typ użytkownika”. Przykładowo, jeśli strona ma być skierowana do młodzieży („Generacji Z” czy „Alfa”), należy uwzględnić potrzeby młodych ludzi i ich sposób recepcji informacji wyszukanych w internecie. W zakończeniu wystąpienia dyrektor biblioteki AGH zapoznał uczestników z wynikami badań przeprowadzonych wśród studentów. Badano, jak wyglądają specjalistyczne bazy danych z punktu widzenia InfoArchu. W opinii studentów dobrze pod tym względem wypada BazEkon.

Drugą prelekcję w tej sesji poprowadził prof. Przemysław Korytkowski, reprezentujący Ośrodek Przetwarzania Informacji – Państwowy Instytut Badawczy oraz Zachodniopomorski Uniwersytet Technologiczny w Szczecinie. Mówca skupił się na zagadnieniu sztucznej inteligencji i jej wpływie na rozwój bibliograficznych baz danych. Zauważył, że AI może wspierać bazy choćby w zakresie automatyzacji pozyskiwania danych z nowych czy archiwalnych publikacji naukowych. W wystąpieniu przedstawiona została też koncepcja informatycznego systemu wyodrębniania cytowań z przypisów dolnych przy pomocy algorytmu przetwarzania języka naturalnego oraz algorytmu widzenia maszynowego. Profesor zauważył, że AI lepiej pracuje – rozpoznaje i odróżnia elementy (np. autor, tytuł, nr strony) – gdy w tekście są klasyczne, dolne przypisy niż wtedy, gdy są one np. w bocznym panelu. Prelegent podkreślił też, że zachodzi bardzo intensywny rozwój sztucznej inteligencji i modele AI wciąż są uczone nowych rzeczy. W założeniach omawianego systemu każdy fragment przypisu powinien zostać odpowiednio otagowany, by został poprawnie rozpoznany przez AI. Gdy AI pracuje na pliku PDF, generuje z artykułów referencje i wszystkie potrzebne podczas wyszukiwania treści, np. liczbę stron czy ważne pojęcia. AI jest wykorzystywana również do czuwania nad poprawnością danych wprowadzonych do baz i w razie potrzeby zajmuje się deduplikacją powielonych rekordów (np. w bazie Polskiej Bibliografii Narodowej większość jest już zdeduplikowana), a także tworzeniem złotych rekordów. Pod koniec wystąpienia prof. Korytkowski podjął temat bazy Ludzie Nauki, której jest współtwórcą. Baza gromadzi informacje o wszystkich naukowcach w polskich instytucjach naukowych. Jest ona wciąż na etapie intensywnego rozwoju i dziś zawiera ponad 200 000 profili naukowców. Obecnie trwają prace nad wyszukiwarką semantyczną, mającą ułatwić intuicyjne przeszukiwanie zawartości bazy przy wykorzystaniu języka naturalnego. Profesor wyraził nadzieję, że Ludzie Nauki staną się szeroko rozpoznawalnym, wiarygodnym źródłem informacji o polskich naukowcach i ich osiągnięciach.



Jako ostatni w sesji trzeciej wystąpił prof. Zbigniew Nawrat, reprezentujący

Fundację Rozwoju Kardiologii im. prof. Zbigniewa Religi na Śląskim Uniwersytecie Medycznym. Jako redaktor czasopisma „Medical Robotics Reports” (MRR) postawił pytania, czy pismo to będzie w przyszłości tworzone przez AI i czy stanie się także czasopismem dla robotów (a jeśli tak, to kiedy). Profesor Nawrat wyraził jasno przekonanie, że dziś poszukiwanie wiedzy w książkach, czasopiśmie zdaje się wielu użytkownikom stratą czasu – możemy ją uzyskać dużo łatwiej i szybciej, zadając odpowiednie pytanie Chatowi GPT. Profesor Nawrat rozważał ten stan rzeczy także w kontekście pracy redaktora czasopisma naukowego, promującego roboty i robotykę na gruncie nowoczesnej medycyny. Prelegent zauważył, że czasy pandemii oraz postępy w dziedzinie informatyki i robotyki sprawiły, że dziś czasopisma mają rację bytu wyłącznie, gdy działają w wersji elektronicznej, najlepiej na otwartej licencji. Odnosząc się do własnego pisma – MRR, mówił o planach poszerzenia grona odbiorców o roboty, bo to one będą adresatami nieodległej przyszłości. Prelegent dodał też, że człowiekowi nie potrzeba dziś nowych kompetencji, bo niedługo nie będzie trzeba wiedzieć ani pamiętać nic, gdyż technologia dostarczy nam danych. Najważniejszą kompetencją we współczesnym świecie jest umiejętność zadawania pytań i ciekawość poznawcza. Zdaniem profesora ważną okazuje się zmiana optyki i zdanie sobie przez naukowców sprawy, że warto pisać artykuły właśnie dla robotów, po to, by one mogły pisać dla innych robotów. Profesor dał wskazówki, jak należy to robić. Przede wszystkim polecenia kierowane do robotów powinny być uporządkowane (warto stosować podziały sekcji – tabele, wykresy, metadane, znaczniki), podane w sposób przejrzysty i jednoznaczny, aby AI mogła je dekodować. Odpowiadając na pytanie zadane w tytule wystąpienia, prelegent zaznaczył, że redagowane przez niego czasopismo już dziś jest pisane przez AI, a w najbliższej przyszłości będzie tworzone także dla AI i robotów.

Na pytanie prof. Korytkowskiego o to, czy można już mówić o tym, że AI nas wyręcza, prof. Nawrat odpowiedział, że ważną rolą człowieka jest teraz odróżnianie tzw. halucynacji AI (przekłamań) od prawdy i że nie można być bezkrytycznym wobec informacji przekazanych przez bota. Sztuczna inteligencja pomaga, ale nie zwalnia z myślenia.

W ramach ciekawostki prelegent poinformował zgromadzonych, że obecnie trwają prace nad tym, by roboty pracujące w medycynie umiały odpowiedzieć na pytanie, dlaczego podjęły jakąś decyzję (np. o wyłączeniu jakiejś funkcji podczas operacji). Roboty są uczone udzielania odpowiedzi, lecz tu rodzi się kolejne pytanie. Co się stanie, gdy roboty same będą umiały zadać pytanie. Pozostaje to na razie bez odpowiedzi.



Moderator sesji, dr Rosiński z Instytutu Badań Literackich PAN, w celu podsumowania tej części zapytał prelegentów, jakie kompetencje są potrzebne do tworzenia nowoczesnych baz danych. Dr Skórka wyraził przekonanie, że najważniejsze jest umiejętne monitorowanie potrzeb rynku oraz pojawiających się nowości i reagowanie na nie – bazy danych, by mieć odbiorców, muszą iść z duchem czasu.

Sesję czwartą poświęcono aspektom prawnym i wyzwaniom polityki otwartego dostępu. Moderowała ją dr Katarzyna Weinper, dyrektor Centrum Informacji Naukowo-Technicznej Politechniki Lubelskiej.

Pierwszą osobą, która została zaproszona do wystąpienia, była Magdalena Szuflińska-Żurawska reprezentująca Bibliotekę Politechniki Gdańskiej. Prelegentka mówiła o Open Knowledge Graphs jako „przykładzie eksploracji komunikacji naukowej w oparciu o integrację danych pochodzących z różnych źródeł”. Prelegentka zwróciła uwagę na fakt, że wyzwaniem dla twórców baz bibliograficznych jest tworzenie adekwatnych powiązań między poszczególnymi elementami dorobku naukowego (linked data). Powstanie specjalistycznych wykresów wiedzy (Open Knowledge Graphs, OKG), które są przykładem integracji danych pochodzących z różnych źródeł (bazy bibliograficzne i bibliometryczne, repozytoria, obiekty cyfrowe) i przedstawienie ich relacji w strukturze hierarchicznej, jest odpowiedzią na zapotrzebowanie środowiska naukowego. Porównując OKG z Google Scholar, prelegentka podkreśliła, że jakość informacji oraz głębokość śledzenia powiązań rekordów okazuje się zdecydowanie większa w OKG. Informacje tam zawarte są pełniejsze i dużo dokładniejsze, dotyczą różnych aspektów, np. finansowania, rodzajów dostępu do publikacji czy ich popularności na podstawie liczby cytowań. W prezentacji zostały omówione najważniejsze przykłady Open Knowledge Graphs (OpenAIRE Knowledge Graph oraz Open Research Knowledge Graph) wraz z ich praktycznym wykorzystaniem przez środowisko naukowe. Podczas wystąpienia prelegentka podzieliła się także refleksją na temat tzw. Deklaracji Barcelońskiej (DB) w sprawie Otwartych Informacji o Badaniach (Barcelona Declaration on Open Research Information), która jest inicjatywą środowisk naukowych na rzecz przekształcenia sposobu tworzenia i wykorzystywania informacji badawczych w celu doprowadzenia ich do otwartości zgodnie z zasadą: „tak otwarte, jak to możliwe, tak jak zamknięte, jak to konieczne”. Prelegentka poinformowała, że do grona sygnatariuszy należy obecnie ponad 40 organizacji naukowych ze świata, lecz żadna z polskich uczelni nie znalazła się w tym gronie. Organizacje, które podpisały DB, dbają o jawność metadanych, informacji o oprogramowaniu, źródła finansowania i zespołów badawczych. Zaletami Open Research Information są także darmowy dostęp, etyczność, dostępność oraz wspieranie współpracy międzynarodowej.



Kolejny referat odtworzony został w formie nagrania. Jego autorem był prof.

Grzegorz

Tylec

z Katolickiego Uniwersytetu Lubelskiego. Mówił on o aspektach prawnoautorskich przy wykorzystaniu dziedzinowych baz danych w procesie szkolenia sztucznej inteligencji. W treści referatu określone zostały ramy prawne dotyczące wykorzystania algorytmów sztucznej inteligencji, zawarte w projekcie rozporządzenia UE – akcie o sztucznej inteligencji. Prof. Tylec poinformował uczestników, że unijny projekt rozporządzenia nie rozstrzyga jednoznacznie konieczności respektowania prawa autorskiego przy wykorzystywaniu danych na potrzeby szkolenia AI, mimo iż zdaje się ona oczywista i że obowiązkiem deweloperów powinno być informowanie o treściach wykorzystywanych do szkolenia sztucznej inteligencji. Prelegent wskazał, że są plany powołania urzędu ds. AI, jednak pojawiają się obawy, że nawet wówczas sprawa praw autorskich w kontekście AI nie zostanie ostatecznie rozwiązana. Mówiąc o aspektach prawnych ewentualnego naruszenia praw autorskich w związku ze szkoleniem algorytmów AI, profesor zwrócił uwagę na problematyczność tego zagadnienia. Otóż zgodnie z podstawą prawną, na mocy licencji CC BY i CC0 autor wyraża zgodę na eksploatację własnych wytworów. Ponadto, zgodnie z ustawą o prawie autorskim, dozwolona jest też niekiedy eksploatacja bez zgody twórcy. Mowa tu o prawie cytowania oraz wykorzystania treści na użytek własny czy naukowy. Płyne stąd wniosek, że autor nie jest w stanie wyłączyć, zastrzec zgody na elektroniczne wykorzystanie własności intelektualnej nawet w celach komercyjnych – tu można bowiem skorzystać z prawa cytatu. Wykorzystanie przez deweloperów systemów AI materiałów zawartych w otwartych dziedzinowych bazach danych (chronionych prawem autorskim) nie stanowi więc naruszenia prawa, gdyż nie ma możliwości prawnej, by dokonać zastrzeżenia. Jednakże według profesora, mimo braku precyzji w rozporządzeniu, AI powinna korzystać z tekstów po uzyskaniu zgody autorów. Byłoby to etycznie poprawne. Prelegent mówił też o konstrukcji „opt out” wyłączającej wyjątek w zakresie nieodpłatnej eksploracji tekstów i danych (*text and data mining* – TDM) do celów komercyjnych przy wykorzystaniu baz danych. Profesor podjął próbę odpowiedzi na pytanie, czy szkolenie AI odbywa się z poszanowaniem TDM i stwierdził, że cyfrowa eksploracja tekstu, czyli analiza informacji w celu zdobycia innych informacji, wykorzystuje zasoby inaczej niż AI. Sztuczna inteligencja kreuje pewne wytwory na podstawie odnalezionych danych, a TDM pełni bardziej odtwórczą rolę.

Wystąpienie profesora zakończyło czwartą sesję.

Sesja piąta, ostatnia podczas tegorocznej konferencji, moderowana przez prof. Jadwigę Woźniak-Kasperek z Uniwersytetu Warszawskiego, została poświęcona znaczeniu bibliograficznych baz danych dla nauki, dydaktyki i praktyki.

Strona 14 z 20

Umowa nr KONF/SP/0420/2023/01

Projekt pt. **IV Konferencja Naukowa BazTech „Bibliograficzne bazy danych: orientacja na przyszłość”**

dofinansowany z programu Ministra Nauki i Szkolnictwa Wyższego:

Dokształcająca Nauka II – Wsparcie konferencji naukowych



Pierwsze wystąpienie dotyczyło znaczenia, jakie mają dla użytkowników

dziedzinowe bazy danych – BazEkon i BazTech. Prelegentkami były dr Anna Kamińska z Biblioteki Politechniki Śląskiej, reprezentująca jednocześnie Instytut Nauk o Kulturze na Uniwersytecie Śląskim, oraz Lidia Derfert-Wolf z ramienia Konsorcjum BazTech. Podczas wystąpienia zaprezentowane zostały wyniki pionierskich

w Polsce badań, dotyczących wpływu dziedzinowych baz na użytkowników. Respondenci zostali poproszeni o ocenę korzyści płynących z korzystania z baz oraz określenie ich wpływu na edukację, pracę naukową i zawodową. Ankieta została skierowana w szczególności do studentów, pracowników naukowych oraz redaktorów i wydawców czasopism korzystających z BazTech i BazEkon. Zdecydowana większość użytkowników wskazała, że bazy te są użyteczne i pomocne, gdyż rozwijają wiedzę, dostarczają informacji o źródłach czy pomagają w projektach. W badaniu padło pytanie o cel, który przyświeca użytkownikom przy wyborze danej bazy. Okazuje się, że ludzie najczęściej sięgają do BazTech i BazEkon przy gromadzeniu materiałów do pracy naukowej, weryfikacji publikacji, a także w celu przygotowania się do zajęć czy sprawdzenia cytowań własnych tekstów. Niektórzy wskazywali też na to, że dzięki bazom mogą śledzić rynek wydawniczy w danej dziedzinie. Wyniki badań pozwoliły twórcom baz na poznanie oczekiwań i kompetencji użytkowników – wielu z nich to specjaliści w danej dziedzinie, nie są to przypadkowe osoby. Ponadto badania wykazały, że współczesne społeczeństwo jest zmęczone i znudzone natłokiem informacji. Dlatego dziedzinowe bazy, które filtrują i dokonują selekcji informacji, są tak pomocnym i ważnym źródłem.

Przedostatnie wystąpienie poprowadziła dr Drabek. Tym razem prelegentka podjęła temat nietrwałości polskich czasopism naukowych. W prezentacji został przedstawiony problem periodyków, które przestają się ukazywać, co często oznacza, że znikają także ich witryny i zgromadzone tam archiwa. Dr Drabek rozpoczęła referat od informacji, że liczba czasopism naukowych we wszystkich dyscyplinach wciąż wzrasta, lecz wiele z nich przechodzi na wydania elektroniczne, a większość nowo powstałych w ogóle nie ma już wersji drukowanej. Jest to związane z łatwością dostępu do tekstów w sieci, a także pozwala na zmniejszenie kosztów związanych z drukiem czy dystrybucją. Z jednej strony pojawia się na rynku wiele nowych czasopism, z drugiej zaś stare zostają zamknięte lub zaprzestają publikowania – stają się nieaktywne. W ostatnich latach z bazy Arianta zostało usuniętych ok. 1000 tytułów. Powodem usunięcia jest najczęściej brak dostępu do strony czy domeny pisma. Często zdarza się też, że w momencie, gdy wydawca zakłada nową stronę internetową, stare wersje plików PDF znikają. Problem stanowią też nieaktualne i/lub niedziałające linki na wciąż istniejącej stronie czasopisma. Prelegentka



zauważyła, że większość polskich czasopism nie ma opracowanej polityki ochrony

danych

i długoterminowej archiwizacji zasobów cyfrowych i że warto się nad tym zagadnieniem pochylić.

W 2020 r. DOAJ zainicjowało projekt JASPER umożliwiający darmową archiwizację cyfrową czasopism. Oferowany jest on w otwartym dostępie wydawcom pism indeksowanych w serwisie DOAJ. Dr Drabek wskazała też na ważną rolę baz danych oraz biblioteki cyfrowej w archiwizacji. Umożliwiają one dostęp do źródła i, zbierając metadane (a czasem i pełne teksty), stają się często jedynym dowodem tego, że dany periodyk w ogóle istniał i zapewniają mu stabilność w sieci. Prelegentka pod koniec wystąpienia wezwała redaktorów i wydawców czasopism, by indeksowali je w bazach i dbali o archiwizację. Podkreśliła, że archiwizowanie jest niezwykle ważne, bo bez niego czasopisma mogą zniknąć (wbrew powszechnie powtarzanemu stwierdzeniu, że w sieci nic nie ginie). Dr Drabek dała też słuchaczom wskazówkę, by w sytuacji, gdy nie można znaleźć archiwalnego tekstu w dziedzinowej bazie danych, spróbowali wyszukać go poprzez Web Archive. Jest to archiwum regularnie uzupełniane przez boty, które automatycznie skanują internet, zbierają dane i archiwizują różnorodne zasoby.

Jako ostatnia prelegentka w sesji piątej wystąpiła, uczestnicząca wcześniej w panelu dyskusyjnym, Dorota Bazan, która mówiła o statystykach wykorzystania zasobów i zadała pytanie, na ile ich prowadzenie to pożyteczne narzędzie diagnostyczne. Statystyki to jedna z metod weryfikacji wykorzystywania zasobów naukowych udostępnianych w różnych systemach. Dają one relatywnie najwięcej informacji i jednocześnie są stosowane w najmniejszym zakresie. Służą m.in. do sprawdzania liczby wejść na stronę i pobrań plików.

To końcowe wystąpienie stanowiło próbę „wskazania metod odczytywania danych statystycznych umożliwiających uzyskanie informacji pomocnych w pracy z serwisami udostępniającymi treści naukowe”. Do analiz wykorzystano dane statystyczne baz prowadzonych w systemach wytworzonych przez ICM UW. Statystyki te pokazują ruch w internecie. W zestawieniach widać, skąd użytkownicy wchodzi do baz czy artykułów, np. z wyszukiwarek, z mediów społecznościowych, ze stron zewnętrznych czy z baz specjalistycznych. Wnioskiem płynącym z analizy zestawień jest wyraźny wzrost wejść bezpośrednich z baz. Według prelegentki jest to znak, że są one coraz lepiej rozpoznawane, a wyszukiwanie jest bardziej specjalistyczne.

Po ostatnim referacie moderująca sesję prof. Woźniak-Kasperek zaprosiła zgromadzonych do krótkiej dyskusji. Podjęto w niej bardziej szczegółowe kwestie dotyczące bazy Ludzie Nauki.



Następnie prof. Woźniak-Kasperek dokonała podsumowania konferencji.

Wskazała m.in. na dwa główne obszary refleksji formułowanych w czasie tego dwudniowego spotkania.

Pierwszy z nich obejmował niejako zagadnienia podstawowe, a mianowicie problematykę budowania, rozwijania

i utrzymywania baz. Drugi natomiast dotyczył kwestii budowania pomostów między bazami a innymi elementami infrastruktury naukowej, integrowania środowisk i dbania o wartości w celu zwiększenia efektywności, użyteczności i wpływu społecznego baz i szerzej pojętej nauki.



3. Wnioski

Wystąpienia podczas czwartej edycji konferencji konsorcjum BazTech uwidoczniły kluczową rolę bibliograficznych baz danych w procesie wyszukiwania, oceny oraz udostępniania informacji naukowej. Pomimo łatwości dostępu do wyszukiwarek takich jak Google Scholar i ich rosnącej popularności, to właśnie bazy danych oferują wysoki poziom merytoryczny, uporządkowaną strukturę metadanych oraz większą wiarygodność indeksowanych treści.

Podczas konferencji zwrócono uwagę na to, że tylko spójna struktura informacji zapewnia ich wysoką użyteczność oraz umożliwia skuteczne wdrażanie automatycznych procesów, takich jak deduplikacja, semantyczne powiązania rekordów czy ocena wiarygodności źródeł. Szczególnego znaczenia nabiera tu koncepcja tzw. złotego rekordu – jednego, pełnego i ustandaryzowanego opisu publikacji, który może być współdzielony pomiędzy różnymi systemami.

Uczestnicy konferencji zidentyfikowali potrzebę zintegrowania rozproszonych baz bibliograficznych, takich jak BazTech, BazEkon i BazHum. Obecna fragmentacja ogranicza zasięg i siłę oddziaływania poszczególnych inicjatyw, podczas gdy ich połączenie mogłoby znacząco zwiększyć rozpoznawalność oraz dostępność polskiego dorobku naukowego. Tego rodzaju integracja sprzyjałaby także ujednoliceniu jakości metadanych i zwiększeniu efektywności prac redakcyjnych.

Pojawiło się również spostrzeżenie, że ważną funkcję wspierającą w pracy z danymi bibliograficznymi może pełnić sztuczna inteligencja, zwłaszcza w zakresie automatycznej kontroli poprawności danych, ekstrakcji cytowań, deduplikacji czy semantycznego wyszukiwania. Warto jednak podkreślić, że algorytmy AI – choć coraz bardziej zaawansowane – nie są wolne od błędów interpretacyjnych („halucynacji”) i powinny działać pod nadzorem merytorycznym człowieka. Tylko takie podejście gwarantuje zachowanie wysokiej jakości i wiarygodności danych.

Otwarta nauka i polityka Open Access wymagają skoordynowanych działań na poziomie krajowym i instytucjonalnym. Kluczowe jest nie tylko promowanie praktyk otwartości, ale także aktywne włączanie się polskich instytucji naukowych w międzynarodowe inicjatywy poprzez podpisywanie i wdrażanie stosownych deklaracji. Równolegle należy dążyć do ujednolicenia stosowanych licencji oraz jednoznacznego uregulowania kwestii prawnych związanych z metodami *text and data mining* (TMD), które obecnie są interpretowane w sposób niespójny w różnych serwisach.

Architektura informacji i projektowanie interfejsów baz powinny uwzględniać potrzeby zróżnicowanych grup odbiorców. Należy pamiętać, że ich zasoby są wykorzystywane zarówno przez młodych badaczy,



studentów, jak i osoby starsze czy użytkowników z niepełnosprawnościami.

Dlatego projektowanie UX musi być oparte na testach użyteczności oraz dostosowane do standardów dostępności, a zmiany powinny być wprowadzane regularnie w oparciu o informacje zwrotne od użytkowników.

Kolejnym istotnym tematem jest zachowanie trwałości zasobów cyfrowych. Brak długoterminowych polityk archiwizacyjnych prowadzi do ryzyka utraty wartościowych materiałów, zwłaszcza w kontekście czasopism elektronicznych o nieustabilizowanej sytuacji wydawniczej.

Rozwój krajowych baz danych wymaga z pewnością pogłębienia dialogu instytucjonalnego oraz zapewnienia stabilnego finansowania. Ważne w tym obszarze jest nawiązanie współpracy z instytucjami, takimi jak Biblioteka Narodowa czy Stowarzyszenie Bibliotekarzy Polskich. Należy dążyć do stworzenia wspólnej strategii na rzecz lepszego wsparcia, promocji i rozwoju krajowych baz danych.

4. Rekomendacje

1. Zaleca się opracowanie i wdrożenie jednolitych standardów tworzenia metadanych, które umożliwią opracowanie tzw. złotego rekordu – ustandaryzowanego, kompletnego opisu publikacji, współdzielonego przez różne systemy. Umożliwi to integrację baz danych oraz automatyzację analiz bibliometrycznych.
2. Rekomenduje się podjęcie działań zmierzających do integracji istniejących dziedzinowych baz danych takich jak BazTech, BazEkon, BazHum. Tego rodzaju integracja zwiększy widoczność polskiego dorobku naukowego, poprawi jakość danych, ich bezpieczeństwo i pozwoli na efektywniejsze wykorzystanie zasobów.
3. Należy rozwijać narzędzia oparte na sztucznej inteligencji, wspomagające automatyczne rozpoznawanie cytowań, kontrolę poprawności metadanych oraz identyfikację i scalanie zduplikowanych rekordów. Jednocześnie konieczne jest zapewnienie stałego nadzoru eksperckiego oraz świadomości ograniczeń tych technologii.
4. Rekomenduje się intensyfikację działań promujących otwartą naukę, w tym aktywne włączanie się instytucji naukowych w międzynarodowe inicjatywy.
5. Interfejsy baz danych powinny być projektowane w sposób inkluzywny, z uwzględnieniem potrzeb różnych grup użytkowników – w tym osób starszych oraz z niepełnosprawnościami. Rekomenduje się prowadzenie regularnych testów użyteczności oraz wdrażanie usprawnień na podstawie informacji zwrotnych.
6. Należy wdrożyć strategię długoterminowej archiwizacji czasopism, z wykorzystaniem dostępnej infrastruktury. Obecnie brak jest powszechnego nawyku korzystania z tych rozwiązań – dlatego konieczne jest podejmowanie działań mających na celu jego wykształcenie i utrwalenie.
7. Rekomenduje się systematyczne gromadzenie i analizowanie danych dotyczących ruchu i zachowań użytkowników w bazach danych. Dane te powinny służyć jako podstawa do podejmowania decyzji rozwojowych oraz promocji zasobów w najbardziej efektywnych kanałach.
8. Ważne jest budowanie współpracy pomiędzy operatorami baz a instytucjami takimi jak Biblioteka Narodowa czy Stowarzyszenie Bibliotekarzy Polskich – np. w zakresie ujednolicania standardów, wspólnego planowania działań i wdrażania strategii. Rekomenduje się opracowanie długofalowej strategii rozwoju i promocji krajowych baz danych oraz zapewnienie stabilnego modelu ich finansowania.