

Sprawozdanie z XIX edycji seminarium z cyklu Digitalizacja

XIX już seminarium z cyklu Digitalizacja, organizowane przez Centrum Promocji Informatyki, odbyło się 26 czerwca 2012 roku w Warszawie. Zatytułowane było „Zarządzanie treścią w bibliotekach cyfrowych – ryzyka prawne, problemy prawa autorskiego”.

Obrady prowadził dr Henryk Hollender, dyrektor Biblioteki Głównej Uczelni Łazarskiego w Warszawie. Referenci, którzy zaprezentowali uczestnikom seminarium swoje wystąpienia, to zarówno teoretycy jak i praktycy zajmujący się informacją naukową, oraz prawnicy.

Jako pierwsza wystąpiła prof. dr hab. Barbara Sosińska-Kalata z Instytutu Informacji Naukowej i Studiów Bibliologicznych Uniwersytetu Warszawskiego. Wykład zatytułowany „Systemy organizacji wiedzy w środowiskach sieciowych” stanowił teoretyczne wprowadzenie w problematykę konferencji. Część wstępna poświęcona była historycznemu spojrzeniu na organizację wiedzy. Autorka sięgnęła do pochodzącego z końca wieku XVII systemu Leibniza, przywołała teorie Peignot’a, Lelewela, Otlea, Deweya, Blissa i Ranganathana. Prezentacja współczesnego podejścia do systemu organizacji wiedzy rozpoczęła się od nawiązania do postulatów H.E. Blissa. Autorka zwróciła uwagę, że „cyfryzacja i globalizacja dostępu do wiedzy” powoduje renesans podejścia Blissowskiego. Uczestnikom zaprezentowała węższe i szersze podejście do organizacji wiedzy jako obszaru badawczego oraz definicje systemu organizacji wiedzy. Według p. profesor, SOW w sensie szerokim to dostęp do wiedzy poprzez m.in.: biblioteki tradycyjne i cyfrowe, muzea i e-muzea, archiwa i e-archiwa, encyklopedie i e-encyklopedie, repozytoria cyfrowych dokumentów i portale wiedzy. Zasadnicza część wystąpienia dotyczyła SOW działających jako sieciowe systemy informacyjne. System organizacji wiedzy obejmować powinien wszystkie rodzaje schematów organizowania informacji i wspierania zarządzania wiedzą (wykazy terminów lub symboli, kartoteki wzorcowe, słowniki haseł przedmiotowych, tezaury, taksonomie organizujące zasoby serwisów internetowych, leksykalne bazy danych, sieci semantyczne). Wszystkie wymienione narzędzia mają umożliwiać pojęciową organizację i dostęp do cyfrowych zasobów wiedzy. Bardzo szczegółowo została omówiona typologia SOW stosowanych w środowisku sieciowym. Na koniec wystąpienia autorka przedstawiła zestaw kryteriów ewaluacji systemów organizacji wiedzy. Podkreśliła, że „decyzja o tym, jaki SOW wybrać spośród bogatej i różnorodnej klasy tych narzędzi, jest jedną z najważniejszych dla każdej kolekcji cyfrowej. System organizacji wiedzy musi być bowiem dostosowany do specyfiki treści zawartych w kolekcji, zakładanych sposobów jej prezentowania oraz zrozumiały dla użytkowników kolekcji”.

Drugie wystąpienie „Budowanie repozytorium instytucjonalnego”, zaprezentowane zostało przez Emilię Karwasińską i Małgorzatę Rychlik z Biblioteki Uniwersyteckiej Uniwersytetu Adama Mickiewicza w Poznaniu.

Autorki, które współtworzą AMUR (Adam Mickiewicz University Repository) podzieliły się z uczestnikami seminarium wiedzą zdobytą w trakcie tworzenia i zarządzania repozytorium. Omówiły główne etapy budowania repozytorium czyli: projektowanie, wdrażanie i rozwój. Podkreśliły, że każdy z nich jest niezwykle ważny. W trakcie projektowania szczególną uwagę należy zwrócić na teoretyczną analizę zagadnienia, zbadać nastawienie środowiska akademickiego do projektu i jego idei, wybrać najlepsze oprogramowanie, określić harmonogram działań oraz kosztorys planowanych prac. Wdrożenie to implementacja i dostosowanie wybranego oprogramowania do potrzeb instytucji oraz tworzenie cyfrowej kolekcji. Omawiając etap wdrożenia autorki zwróciły uwagę na aspekt zarządzania użytkownikami repozytorium obejmujący: rejestrację użytkowników uprawnionych do deponowania prac, przypisywanie nowych użytkowników do odpowiednich grup, nadawanie

uprawnień depozytariuszom, archiwizowanie prac przez uprawnionych użytkowników). Podkreślili również znaczenie przemyślanego doboru materiałów do budowanej kolekcji.

Dużą część wystąpienia poświęcona była kwestiom prawa autorskiego. Autorki odpowiadały na pytania: kiedy autor może zamieścić własne publikacje w repozytorium, na jakich zasadach może je udostępniać, podały również przykłady zastosowania licencji Creative Commons. Jednym z przykładów podanych przez autorki była zawartość licencji niewyłącznej, którą autor zatwierdza w procesie deponowania. Etap rozwoju repozytorium to między innymi jego promocja (w tym rejestracja w katalogach Open Access). Na koniec autorki zwróciły uwagę, że w odróżnieniu od trendów światowych (w wielu państwach wspiera się tworzenie repozytoriów poprzez opracowywanie i wdrażanie ogólnokrajowych programów służących ich rozwojowi) w Polsce nie wypracowano jeszcze strategii w tym zakresie.

Kolejne wystąpienie zatytułowane „Proces doboru dokumentów do biblioteki cyfrowej i kontekst prawny: casus Biblioteki Cyfrowej Polskiego Instytutu Antropologii” zaprezentowane przez Joannę Koźmińską z Polskiego Instytutu Antropologii oraz Instytutu Etnologii i Antropologii Kulturowej Uniwersytetu Warszawskiego, również miało charakter praktyczny. Referentka przedstawiła zebrany bibliotekę cyfrową Polskiego Instytutu Antropologii. Na jej utworzenie duży wpływ wywarł fakt, że jeszcze w roku 2008 (kiedy istniało już w Polsce wiele bibliotek cyfrowych), zawartość polskich treści etnograficznych w ich zasobach była znikoma. Tworzenie BC PIA rozpoczęło od określenia docelowej grupy użytkowników oraz zdefiniowania zawartości kolekcji. Ustalono, że powstanie cyfrowa biblioteka dziedzinowa gromadząca teksty z zakresu etnologii i antropologii kulturowej. Podstawą zasobu miała się stać zawartość trzech najważniejszych czasopism etnograficznych ukazujących się w Polsce po drugiej wojnie światowej. Tytuły tych czasopism to: Lud, Etnografia Polska, Polska Sztuka Ludowa i Konteksty. Ich zawartość to około 4500 artykułów. Indywidualnym rozwiązaniem zastosowanym w PIA jest to, iż każdy artykuł (podobnie jak w zagranicznych bazach danych) traktowany jest jak odrębna publikacja. Zawartość BC PIA nie będzie się ograniczała tylko do tekstów artykułów z czasopism; planuje się także publikacje prac zbiorowych oraz materiałów konferencyjnych. Plany rozwoju BC PIA to: poszerzenie kolekcji książkowej o Studia Ethnologica, zamieszczenie tekstów klasyków antropologii zachodniej, słowników i podstawowych prac metodologicznych i teoretycznych, a także współpraca z Bibliografią Etnografii Polskiej. Największym problemem Biblioteki Cyfrowej PIA jest pozyskiwanie praw autorskich. Zgodnie z Ustawą o prawie autorskim, twórcy bibliotek cyfrowych muszą uzyskać zgodę wydawców i autorów publikacji na umieszczenie ich tekstów w Internecie. Twórcy BC PIA musieli więc dotrzeć do ok. 2 000 autorów, których prace zamieszczone są w zdigitalizowanych czasopismach. Dzięki ustaleniu zasad docierania do autorów i spadkobierców (pomoc prawnika, utworzenie bazy danych o autorach), udało się udostępnić połowę z 4590 zdigitalizowanych tekstów. 310 autorów wyraziło zgodę na publikację ich tekstów w BC PIA. Oznacza to, że na stronie BC można w tej chwili czytać 2121 artykułów. Pozostałe, niestety są niedostępne do czasu uzyskania praw autorskich do ich publikacji.

Autorem kolejnego wystąpienia zatytułowanego „Indeksowanie tekstów w zasobach bibliotek cyfrowych” był dr Piotr Malak z Instytutu Informacji Naukowej i Bibliologii Uniwersytetu Mikołaja Kopernika w Toruniu. Rozpoczął swoje wystąpienie od przypomnienia, że główne zadania bibliotek cyfrowych są identyczne z zadaniami bibliotek tradycyjnych (gromadzenie, opracowanie i udostępnianie zbiorów). Proces indeksowania, który ma miejsce w obu typach bibliotek może być tradycyjny lub zautomatyzowany. Słowne charakterystyki dokumentów przeprowadzane metodą tradycyjną powstają w wyniku opracowania rzeczowego przez indeksatorów. W indeksowaniu zautomatyzowanym (dokumentów cyfrowych) istnieje możliwość automatycznego uwzględniania treści dokumentu w trakcie indeksowania. Dostęp do pełnej treści dokumentu przy pomocy narzędzi wyszukiwawczych, pozwala nie tylko na przeglądanie metainformacji, ale i treści dokumentów. Indeksowanie pełnotekstowe pozwala użytkownikowi samodzielnie wyszukiwać dokumenty

z interesującą go treścią. Odbywa się to poprzez zdefiniowanie słów lub wyrażeń kluczowych. Wadą wyszukiwania zautomatyzowanego jest niższa niż w przypadku indeksowania tradycyjnego precyzja i dokładność. Referent scharakteryzował najbardziej popularne, wykorzystywane w bibliotekach oprogramowanie, czyli dLibre. Omówił również działanie Federacji Bibliotek Cyfrowych. Duża część wystąpienia poświęcona była formatom plików stosowanych w księgozbiorach cyfrowych i wynikającym z tego problemom z dostępem do treści dokumentów. Szczególnie zajął się formatami plików DjVu oraz PDF oraz metodami i problemami konwersji plików do formatu tekstowego. W końcowej części wystąpienia omówione zostały tagi (folksonomie), jako sposób społecznego klasyfikowania treści dokumentów dostępnych w sieci. Podsumowując, autor stwierdził, że wyszukiwarki internetowe przyzwyczyły użytkowników do nowego sposobu wyszukiwania informacji i ma to wpływ na oczekiwania w stosunku do systemów dostarczających treści cyfrowe.

„Głęboki Internet” a zasoby bibliotek cyfrowych – to temat wystąpienia Lidii Derfert-Wolf z Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy. Referentka zaczęła swoje wystąpienie od podania definicji zjawiska „głębokiego Internetu”. Terminem tym określa się zasoby sieci, które nie są indeksowane przez wyszukiwarki (w odróżnieniu od indeksowanej „sieci płytkiej”). Wg badań zasoby głębokie są ponad 400 razy większe niż zasoby płytke. Autorka przytoczyła kilka danych liczbowych: do ok. 95% zasobów ukrytych, użytkownicy mają bezpłatny dostęp; połowa z nich to tematyczne bazy danych, zasoby głębokiego Internetu odwiedzane są przez użytkowników o 50% częściej niż zasoby płytke. Wymieniła następujące rodzaje informacji ukrytych w sieci niewidzialnej: zawartość publicznie dostępnych baz danych, strony i pliki nietekstowe, w formatach innych niż HTML, strony, do których nie prowadzą odsyłacze z innych witryn, strony zbyt „młode” (niedawno powstałe), strony wyłączone z procesu indeksacji przez twórców, zasoby dostępne na hasło, strony i pliki tworzone dynamicznie lub w czasie rzeczywistym, zasoby dostępne w ograniczonym czasie i zasoby dostępne dla określonego terytorium. Wg autorki wystąpienia, biblioteki cyfrowe są przykładem „głębokich” zasobów sieci z racji tego, że najczęściej funkcjonują w postaci baz danych i zawierają obiekty w różnych formatach. Udostępnianie zasobów to zasadniczy cel tworzenia bibliotek cyfrowych, dlatego autorka zwraca uwagę na fakt, jak ważne jest dotarcie do tych zasobów przez użytkownika czyli zwiększenie ich widoczności w sieci. Lidia Derfert-Wolf tak jak jej przedmówca odniosła się do rodzaju plików udostępnianych w bibliotekach cyfrowych, ale z innej perspektywy. Zwróciła uwagę, aby albo unikać formatów dokumentów tekstowych, ponieważ nie są one indeksowane przez roboty wyszukiwarek, albo prezentować treść w kilku formatach; podkreśliła jak ważne jest stosowanie techniki rozpoznawania pisma (OCR) przy skanowaniu dokumentu. Mówiąc o „wyszukiwalności” plików graficznych z kolekcji cyfrowych zauważyła, że staną się one „widoczne” dla użytkownika tylko wtedy, gdy zastosujemy do ich opisu rozbudowane metadane; podobnie rzecz się ma z plikami audio i wideo. Tylko dzięki metadanom będą one indeksowane przez roboty. Wg Lidii Derfert-Wolf sposobem na uzyskiwanie lepszych wyników wyszukiwania jest rozbudowywanie metadanych opisowych (z czym jest problem w bibliotekach cyfrowych). Dane są niespójne, rozbieżne a pola niejednolicie wypełnione, na co autorka zwracała już kilkakrotnie uwagę w innych swoich wystąpieniach. Autorka wskazała również na standardowe protokoły i formaty zapisu danych oraz zacytowała założenia SEO (Search engine optimization – pozycjonowanie stron internetowych) – są to m. in.: odpowiedni dobór słów kluczowych, poprawność techniczna strony, mapa strony, dostępność danych w różnych formatach, wzajemne linkowanie. Na koniec zwróciła uwagę na rolę działań promocyjnych w informowaniu o zasobach biblioteki cyfrowej oraz na to, jak ważne są szkolenia bibliotekarzy cyfrowych i informatyków.

Recepta na zwiększenie widoczności kolekcji cyfrowych w sieci? Bądź tam, gdzie są twoi klienci!

Podsumowaniem obrad była dyskusja panelowa „Problemy prawno-autorskie w bibliotekach cyfrowych”, moderowana przez doktora Henryka Hollendra.

Wystąpienia na temat prawa autorskiego przedstawili: mecenas Piotr Łada, prowadzący kancelarię adwokacką w Warszawie; doktor Zofia Zawadzka z Zakładu Prawa Własności Intelektualnej na Wydziale Prawa Uniwersytetu w Białymstoku i Barbara Szczepańska – kierująca biblioteką i zasobami informacyjnymi w kancelarii prawnej Hogan Lovells w Warszawie. Problemy naświetlane w trakcie wystąpienia mecenasa Piotra Łady dotyczyły: praw autorskich i majątkowych, prawnych podstaw digitalizacji, autorskich praw osobistych, praw pokrewnych. Mecenas Piotr Łada zwrócił uwagę na problem naruszania praw autorskich i ryzyka prawnego w zakresie digitalizacji. Odwoływał się w tym zakresie do konkretnych zapisów w Ustawie z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych. Doktor Zofia Zawadzka w swoim wystąpieniu zajęła się szczegółową analizą artykułu 28 Ustawy, a w szczególności: problemem instytucji dozwolonego użytku na rzecz bibliotek, nieodpłatnym udostępnianiem egzemplarzy utworów rozpowszechnionych, zwielokrotnianiem rozpowszechnianych utworów, udostępnianiem za pośrednictwem terminali. Podkreśliła, że w praktyce biblioteki cyfrowe najczęściej udostępniają utwory, co do których wygasły już autorskie prawa majątkowe. Wystąpienie Barbary Szczepańskiej również dotyczyło artykułu 28 Ustawy i w jego kontekście praw, jakie mają biblioteki, archiwa i szkoły oraz artykułu 27. Podsumowując swoją wypowiedź autorka wskazała na istotne w sensie prawa autorskiego różnice pomiędzy biblioteką cyfrową a repozytorium. Biblioteka cyfrowa gromadzi i udostępnia zasoby zdigitalizowane, o różnym statusie prawno-autorskim; natomiast w repozytorium gromadzi się obiekty powstałe w formie cyfrowej, przekazywane tam przez autora. Na autorze więc spoczywa odpowiedzialność za oznaczenie utworu i jego status prawny.

Po wystąpieniach osób biorących udział w panelu wywiązała się ożywiona dyskusja. Uczestnicy zadawali szereg pytań z prośbą o interpretację i komentarz. Najwięcej pytań dotyczyło artykułu 28 ustawy.

Program seminarium i dobór referentów oraz prowadzącego obrady były bardzo dobrze przemyślane. Uczestnicy mogli wysłuchać zarówno głosów cenionych teoretyków jak i praktyków, na co dzień zmagających się z wieloma problemami dotyczącymi zarówno bibliotek cyfrowych, jak i repozytoriów. Opinie zaproszonych prawników rozwiąły wiele wątpliwości, z którymi bibliotekarze stykają się każdego dnia. Żywa dyskusja potwierdziła, że temat był interesujący, a potrzeba zarówno wymiany informacji jak i doświadczeń związanych z bibliotekami cyfrowymi, ogromna.