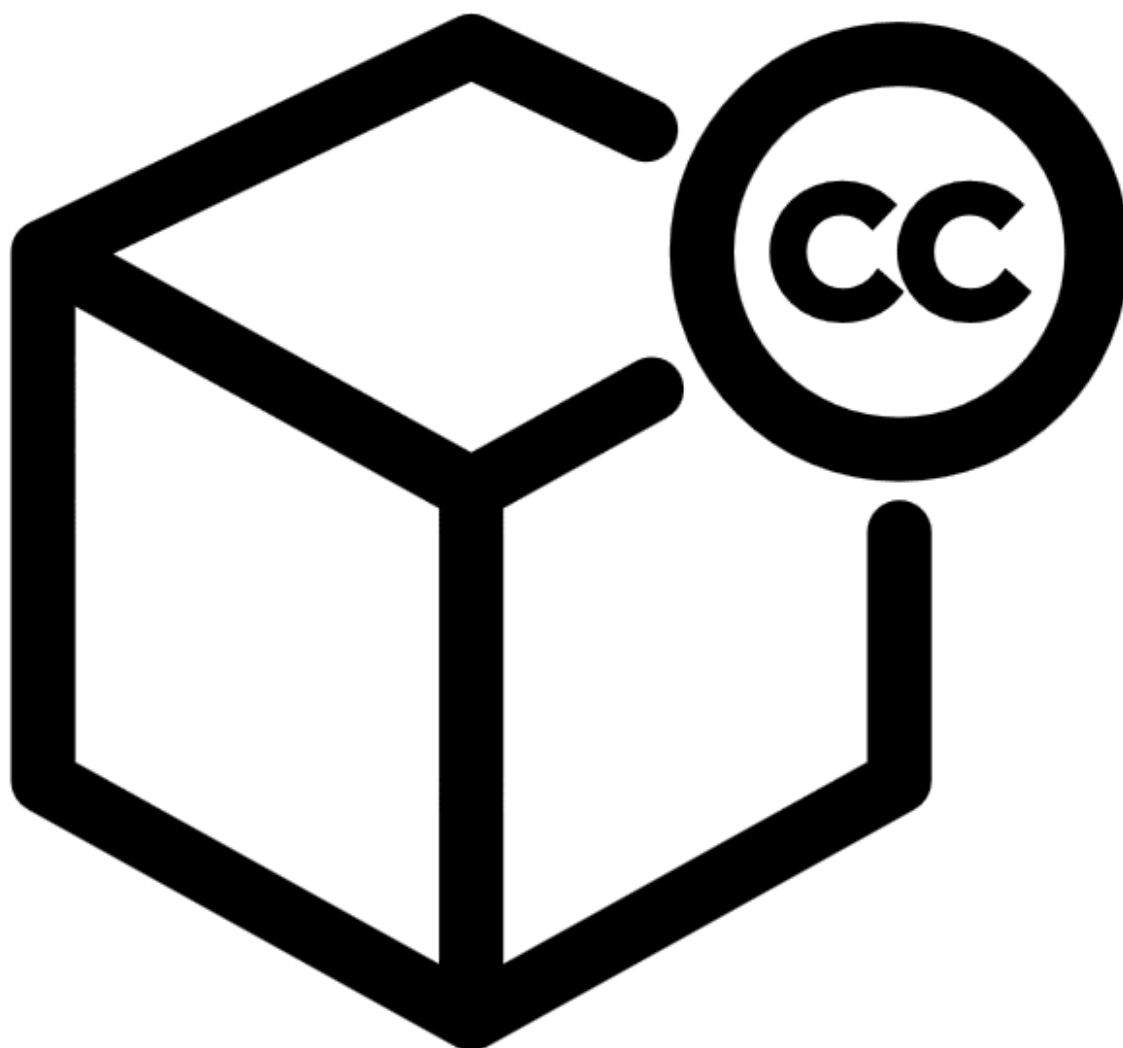


SKALA I SPOSOBY WYKORZYSTANIA LICENCJI CREATIVE COMMONS W POLSCE.



CENTRUM
CYFROWE

projekt:polska®

**SKALA I SPOSOBY
WYKORZYSTANIA LICENCJI
CREATIVE COMMONS
W POLSCE.**

**ANALIZA WYKORZYSTANIA
LICENCJI NA STRONACH
WWW.**

Skala i sposoby wykorzystania licencji Creative Commons w Polsce. Analiza wykorzystania licencji na stronach WWW.

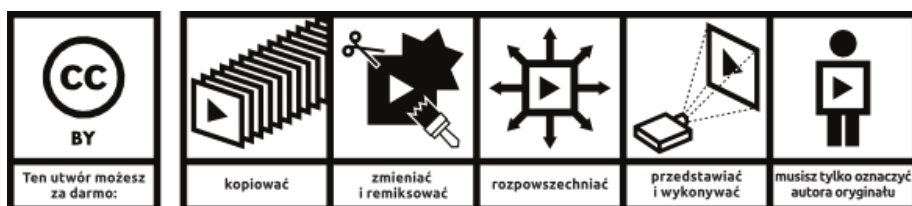
Autorzy: dr Alek Tarkwoski, Ewa Majdecka



Grafika na okładce: [Five by Five](https://thenounproject.com/term/creative-commons-data/38250/), domena publiczna, źródło: <https://thenounproject.com/term/creative-commons-data/38250/>

Warszawa 2015

Analiza jest dostępna na licencji Creative Commons Uznanie autorstwa 3.0 Polska. Pewne prawa zastrzeżone na rzecz autorów oraz Centrum Cyfrowego Projekt: Polska. Zezwala się na dowolne wykorzystanie treści - pod warunkiem zachowania niniejszej informacji licencyjnej i wskazania autorów oraz Centrum Cyfrowego Projekt: Polska jako właścicieli praw do tekstu. Treść licencji jest dostępna na stronie <http://creativecommons.org/licenses/by/3.0/pl/>.



Publikacja jest dostępna w sieci pod adresem:

<http://centrumcyfrowe.pl/czytelnia/skala-i-sposoby-wykorzystania-licencji-creative-commons-w-polsce-analiza-wykorzystania-licencji-na-stronach-www/>

SPIS TREŚCI

Wprowadzenie.....	6
Przegląd literatury.....	7
Metodologia	11
Skala wykorzystania licencji CC w polskim internecie	13
Sposoby opisywania treści na licencjach CC.....	14
Strony internetowe udostępnione na jednej z licencji Creative Commons	15
Strony internetowe z treściami udostępnianymi na licencjach Creative Commons	19
Systematyczne czy incydentalne wykorzystywanie licencji Creative Commons	25
Wnioski i rekomendacje	26
Bibliografia	27

WPROWADZENIE

Prowadząc projekt Creative Commons Polska poświęcamy dużo uwagi edukowaniu twórców różnorodnych treści odnośnie zasad ich udostępniania. Uczymy, że aby stosować licencje Creative Commons (CC) należy także poprawnie oznaczać licencjonowane treści. Szczególnie, że opisanie treści jako licencjonowanej jest jedynym formalnym warunkiem użycia licencji CC.

W 2013 roku przeprowadziliśmy eksperymentalne indeksowanie zasobów polskiego internetu. Naszym podstawowym celem było zmierzenie skali wykorzystania licencji Creative Commons. Liczba treści na licencjach CC jest podstawową miarą rozwoju ruchu wolnej kultury. W rezultacie uzyskaliśmy też bogatą próbę stron udostępnianych na licencji CC bądź zawierających treści na tej licencji. Pozwala ona na analizę sposobów i poprawności informacji prawnoautorskich dotyczących licencji CC.

Mimo upływu czasu od chwili realizacji badania jesteśmy przekonani, że dane badawcze dostarczają przydatnej wiedzy na temat stosowania licencji CC w Polsce. Jego podstawową zaletą jest oparcie się na empirycznej analizie wykorzystania licencji CC, przeprowadzonej na dużej próbie stron WWW.

PRZEGLĄD LITERATURY

Mimo znaczenia pomiaru wykorzystania licencji CC jako kluczowego wskaźnika rozwoju wolnego licencjonowania, badania tego rodzaju są prowadzone sporadycznie.

Najważniejszym punktem odniesienia są badania przeprowadzone w 2007 roku przez Giorgosa Cheliotisa z Participatory Media Lab¹. Wykorzystywane źródła danych to wyszukiwarki Yahoo! i Google oraz serwisy specjalistyczne (ograniczające się do jednego lub kilku rodzajów medium), np. Flickr².

Najważniejsze wnioski z tych badań koncentrują się wokół bardzo zróżnicowanych szacunków liczb obiektów wykorzystujących CC. Istotną metodologiczną obserwacją są jednak zbliżone wyniki z wyszukiwarek Yahoo! i Google. Wnioski związane z samymi licencjami to chociażby informacja, że 70% licencji CC posiada warunek użycie niekomercyjne. Popularne było również użycie na tych samych warunkach, które związane było z 50% obiektów na licencji CC. Zdecydowanie niepopularna była licencja uznanie autorstwa - na tych samych warunkach - wykorzystywana choćby przez Wikipedię. Polska, w porównaniu do innych krajów posiadała zdecydowanie większy udział licencji uznanie autorstwa - użycie niekomercyjne - na tych samych warunkach. Należy jednak zwrócić uwagę, że próbki z badanych krajów były nierównomierne.

Dzieląc licencje CC na trzy grupy (liberalne (BY & BY-SA), umiarkowane (BY-ND & BY-NC) i restrykcyjne (BY-NC-SA & BY-NC-ND)), Polska (a raczej polska jurysdykcja) jest jednym z krajów, w którym stosuje się najwięcej licencji restrykcyjnych (choć taka tendencja istnieje na całym świecie - szczególnie w jurysdykcjach, gdzie jest powyżej 100.000 obiektów na licencjach CC)³. Licencje umiarkowane były najmniej popularne. Trzeba przy tym zaznaczyć, że zaproponowany podział jest dość kontrowersyjny, szczególnie jeśli chodzi o zakwalifikowanie licencji z warunkiem “Bez utworów zależnych” (ND) do kategorii licencji umiarkowanych.

Metodologia obrana przez Cheliotisa posiada pewne ograniczenia, na które zwraca uwagę sam autor. Pobieranie danych z wyszukiwarek internetowych powoduje pominięcie części stron WWW, na których były umieszczane utwory na licencjach

¹ Giorgos Cheliotis (2007). Creative Commons Statistics from the CC-Monitor Project. <https://hoikoinoi.files.wordpress.com/2007/07/cc-monitor-findings-issummit.pdf>, dostęp 12.03.2015.

² Ibidem. Uproszczony model pozyskiwania danych dostępny na slajdzie 6.

³ Ibidem, slajd 14-15.

Creative Commons. W tym sensie wyniki można odnosić jedynie do zebranych danych, a nie całego internetu.

Odpowiedzi na pytania (Jak wielu autorów używa licencji Creative Commons? Kim są? Które licencje wybierają? Co ma wpływ na ten wybór? Jak porównywać jurysdykcje?), które zespół Cheliotisa zadał sobie przed realizacją badania, są zatem tylko częściowe.

Brak uporządkowanej wiedzy na temat statystyk wykorzystania licencji był przyczynkiem do stworzenia w 2008 roku przez organizację Creative Commons projektu CC Metrics Project. Celem projektu było uporządkowanie metodologii, zebranie i udostępnienie istniejących danych oraz wsparcie dla środowiska badawczego zainteresowanego tym tematem.

CC Metrics zakłada dwa podstawowe źródła danych. Pierwszym są dane o liczbie odnośników do stron z opisem licencji CC (tzw. linkback data). Zliczanie tej statystyki jest możliwe dzięki temu, że CC utrzymuje tylko jedną stronę wzorcową dla każdej licencji, a postanowienia licencji wymagają zamieszczania - tam gdzie to możliwe - odnośników do tych stron. Statystyki te opierają się jednak na wtórnych danych udostępnianych przez wyszukiwarki Google oraz Yahoo, podobnie jak w badaniach Cheliotisa, i charakteryzują się dużą zmiennością i poziomem niepewności. Drugim źródłem są statystyki wykorzystania licencji poprzez duże platformy internetowe⁴ - przede wszystkim serwis Flickr, który jako jeden z pierwszych dużych serwisów komercyjnych umożliwił użytkownikom licencjonowanie treści na licencjach CC.

Wytypowanie jako źródła danych 20 stron internetowych będących repozytoriami treści na licencjach Creative Commons przyczyniło się do kolejnych metodologicznych utrudnień. Nie każda ze stron udostępniała dane potrzebne w projekcie. Część z nich badacze otrzymali dzięki osobistym kontaktom. W tym wypadku również nie można mówić o reprezentatywnym obrazie wykorzystania licencji Creative Commons, gdy dostęp do części danych był niemożliwy.

Projekt CC Metrics jest obecnie nieaktywny, a najnowsze informacje dotyczą planów na 2012 i 2013 rok. Nie działa również serwis CC Monitor, który z założenia miał w sposób automatyczny zbierać dane o wykorzystaniu licencji.

Creative Commons równocześnie zainicjowało projekt CC Case Studies, którego celem była jakościowa analiza rozwoju otwartego licencjonowania - poprzez studia przypadków. Metoda ta jest metodologicznie dobrym rozwiązaniem problemu jakości danych statystycznych - jednak projekt CC Case Studies nigdy nie miał rygoru badawczego i należy go traktować raczej jako formę promocji działań organizacji.

⁴ <https://wiki.creativecommons.org/wiki/Research>, dostęp: 17.07.2015.

Creative Commons rozpoczęło wówczas także prace nad projektem ccREL (Creative Commons Rights Expression Language⁵), którego celem było wypracowanie standardu metadanych dla licencji Creative Commons. Wykorzystując standardy prezentacji danych RDFa oraz XMP, ccREL pozwala opisywać licencjonowane media w sposób pozwalający na odczyt maszynowy. Teoretycznie pozwoliłoby to monitorować skalę wykorzystania licencji, nikt jednak tego nie robi. W oparciu o metadane CC utworzyło natomiast wyszukiwarkę treści na licencjach CC.

Główne problemy badań Creative Commons i zespołu Cheliotisa polegają przede wszystkim na doborze próby, który pozwalał na wyciąganie wniosków dotyczących tylko niej samej. Bardzo nieprecyzyjnie można również określić, jak dana próba ma się do całego internetu. Wyniki są zatem pewnymi wskazówkami bez gwarancji powodzenia na innej próbie.

W 2014 roku Creative Commons rozpoczęło na nowo pomiar wykorzystania licencji, w ramach projektu State of the Commons⁶. Nowe statystyki skali wykorzystania licencji opierają się na bazach danych Google (Google cache). Dane dotyczące rozkładu licencjonowania pomiędzy państwami są oparte na statystykach geolokalizacji użytkowników korzystających z narzędzia wyboru licencji (licence chooser) na stronie CC. Drugim źródłem są dane z ośmiu serwisów stosujących licencje CC i mających ponad milion obiektów na tych licencjach.

W każdym wypadku brak jednak pewności co do kompletności danych. W przypadku danych Google brak jawnych informacji na temat metodologii ich zbierania. W przypadku serwisów internetowych brak pewności, że wszystkie zostały zidentyfikowane. Arbitralne jest również ograniczenie do serwisów z ponad milionem licencjonowanych obiektów.

Ponadto, uwzględniając tempo zmian w internecie opublikowane dane szybko stają się nieaktualne. Zbierane statystyki są raczej dokumentacją określonych, minionych momentów, a porównane ze sobą mogą wskazywać na kierunek rozwoju korzystania z licencji Creative Commons. Z pewnością jednak brak metodologii i narzędzi pomiaru pozwalających stale i precyzyjnie monitorować rozwój wykorzystania licencji CC.

W żadnej z analiz nie znaleźliśmy jednak informacji o korzystaniu z web crawlerów - robotów automatycznie indeksujących i archiwizujących strony WWW. Podjęte

⁵ <https://creativecommons.org/ns>, dostęp: 19.07.2015.

⁶ Creative Commons (2014). State of the Commons: Notes and additional resources. <https://github.com/creativecommons/stateofthe/blob/master/data/notes.md>

przez nas badanie należy więc traktować jako nowatorską próbę wypracowania nowej metodologii pomiaru wykorzystania licencji CC.

W przywołanych tu badaniach istotne jest również ograniczenie analiz wyłącznie do skali wykorzystania licencji. Wybór tego wskaźnika ma zapewne charakter pragmatyczny i wynika z ogólnych trudności pomiaru obiegów treści z wykorzystaniem wolnych licencji. Można jednak przyjąć alternatywne wskaźniki, takie jak na przykład liczba osób i instytucji wykorzystujących licencje czy też liczba użyć licencjonowanych treści (co jest szczególnie trudne do zmierzenia). Pojedyncza miara, jaką jest liczba licencjonowanych obietków, daje nam tylko najbardziej podstawowe spojrzenie na licencjonowanie CC.

Brakuje też badań i refleksji nad tym, jak te treści są opisywane oraz czy przestrzegane są warunki licencji, na której treść została udostępniona. Tymczasem są to równie istotne czynniki związane z wykorzystywaniem licencji, które powinny być analizowane. Nasze badanie ma za zadanie wypełnienie tej luki.

METODOLOGIA

Analiza została przeprowadzona w oparciu o listę stron WWW zawierających frazę “Creative Commons”, zindeksowanych przez robota internetowego (web crawler) w październiku-listopadzie 2013 roku. Zindeksowane zostało 19,6 milionów stron WWW (pojedynczych dokumentów HTML), w ramach 735 tysięcy domen internetowych, które zostały zidentyfikowane jako polskojęzyczne.

Należy podkreślić, że w tego rodzaju badaniu jest niemożliwe zindeksować pełen zbiór stron, nie ma też możliwości oceny, jaki procent tego zbioru został zebrany przez robota internetowego. Zindeksowane przez nas strony traktujemy więc jedynie jako próbę wszystkich stron polskiego internetu. Istnieją też trudności z precyzyjnym zidentyfikowaniem tego zbioru, które można definiować na podstawie domen (tych, które znajdują się w domenie krajowej .pl) lub na podstawie wykorzystanego języka. W obu wypadkach w zbiorze znajdą się jednak również serwisy zagraniczne - kategoria “polskiego zasobu WWW” ma więc nieostre granice.

rodzaj indeksu	liczba
dokumenty polskojęzyczne	19602144
dokumenty polskojęzyczne z frazą “Creative Commons”	243544
dokumenty polskojęzyczne z frazą “Creative Commons” z linkami do licencji	132812
domeny polskojęzyczne	734866
domeny polskojęzyczne z frazą “Creative Commons”	3532
domeny polskojęzyczne z frazą “Creative Commons” z linkami do licencji	1937

Z listy 1937 domen z linkami do licencji CC analizowaliśmy co trzecią domenę, by osiągnąć liczbę 100 domen z treściami udostępnianymi na licencji Creative Commons oraz 100 domen na jednej z licencji Creative Commons. Ostatecznie przeanalizowaliśmy 722 domeny, w tym 115 domen na CC oraz 103 z treściami na CC. Analiza pokazała dużą zmienność i ulotność treści w internecie. Część stron WWW z próby już nie istniała, część znikła w trakcie analizy. To czynnik, który także utrudnia precyzyjny pomiar wykorzystania licencji CC.

Analiza polegała na odpowiedzi na następujące pytania:

- czy strona internetowa działa?;
- czy strona internetowa jest udostępniana jako całość na jednej z licencji Creative Commons?;
- jeżeli tak, jak ta licencja jest opisana?;
- czy jest link do podstrony, gdzie licencja jest opisana;
- czy na stronie są udostępniane treści na licencjach Creative Commons?;
- jeżeli tak, to jaki jest to rodzaj treści?;
- czy utwór jest własny, czy pobrany?;
- czy na stronie znajduje się artykuł o licencjach Creative Commons?;
- jeżeli tak, to jaki jest link do artykułu o Creative Commons;
- jakie licencje Creative Commons są wykorzystywane na stronie WWW licencje Creative Commons;
- ewentualne dodatkowe informacje.

SKALA WYKORZYSTANIA LICENCJI CC W POLSKIM INTERNECIE

Podstawowym wynikiem naszego badania były dane o skali wykorzystania licencji CC w polskim internecie. Jak wspomnieliśmy w nocie metodologicznej, ograniczeniem były trudności ze zdefiniowaniem pełnego zbioru polskich stron WWW - co uniemożliwia podanie faktycznej liczby polskich stron (lub ogólniej treści) na licencji CC. Operując jednak bardzo dużą próbą 19,6 milionów stron możemy z dużym prawdopodobieństwem oszacować procentowy udział takich stron. 1,25% stron zawiera frazę "Creative Commons", a odnośnik do stron z pełną treścią jednej z licencji CC zawiera 0,68% stron. Analogicznie frazę zawiera 0,48% domen, a odnośnik do licencji 0,22% domen. (Różnica w wynikach wiąże się z tym, że czasem jedna domena zawiera więcej niż jedną stronę na licencji CC). **Przyjmujemy umownie, zaokrąglając wyniki, że na licencjach CC jest udostępnianych około 1% wszystkich treści online.**

Można uznać, że udział treści na licencjach CC w treściach polskiego internetu jest marginalny. Ale można też uznać, że stosunkowo duża grupa twórców stron podejmuje świadomą decyzję, by udostępniać treści na otwartych zasadach lub wykorzystać treści dostępne na licencji CC. Biorąc pod uwagę skalę treści dostępnych online, nawet 1% tych treści to znaczący zasób. Należy też pamiętać, że zgodnie z zasadami prawa autorskiego treści są obejmowane ochroną automatycznie po ich "ustaleniu". Tym istotniejsza jest świadoma decyzja posiadacza praw do nich, by udostępnić swoje zasady na innych zasadach.

SPOSOBY OPISYWANIA TREŚCI NA LICENCJACH CC

Wzmianka o licencjach CC może się pojawić na stronie internetowej w trzech wypadkach:

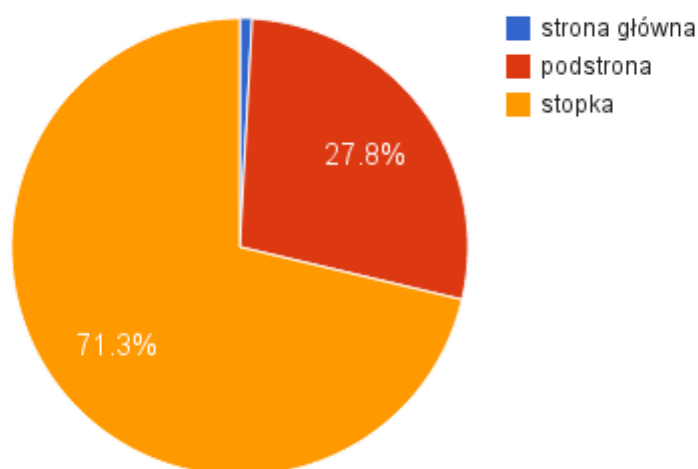
1. treść na stronie jest udostępniona na jednej z licencji Creative Commons;
2. strona udostępnia własne lub cudze utwory na licencjach CC; lub
3. strona wspomina o licencjach, nie zawiera jednak licencjonowanych treści.

Trzeci przypadek jest dla nas nieistotny (oraz marginalny w analizie). Naszą analizę prowadziliśmy według rozróżnienia na dwa pierwsze przypadki.

STRONY INTERNETOWE UDOSTĘPNIONE NA JEDNEJ Z LICENCJI CREATIVE COMMONS

Podstawowym wymogiem przy licencjonowaniu treści na licencji CC jest umieszczenie informacji licencyjnej. Sprawdziliśmy więc sposób prezentowania tych informacji na stronie. 70% domen stosujących licencje CC umieszcza informacje o wybranej licencji Creative Commons w stopce strony. Wydaje się, że to najlepsze miejsce do skutecznego informowania użytkowników o prawach, na których mogą oni wykorzystać treść strony. Posiadają oni bowiem tę informację z poziomu każdej strony serwisu. Jednocześnie pozwala to na lepsze indeksowanie w wyszukiwarkach w kontekście hasła “Creative Commons” i tematu związanego z domeną (np. przy wyszukiwaniu “ogród creative commons”). 30% stron umieszcza te informacje na dedykowanej podstronie - zaletą takiego rozwiązania jest z kolei możliwość umieszczenia obszernych wyjaśnień. Jedynie w pojedynczych przypadkach informacje znajdowały się w innym miejscu - np. gdzieś na stronie głównej.

**Miejsce informacji o umieszczeniu strony na
licencji Creative Commons [n=116]**

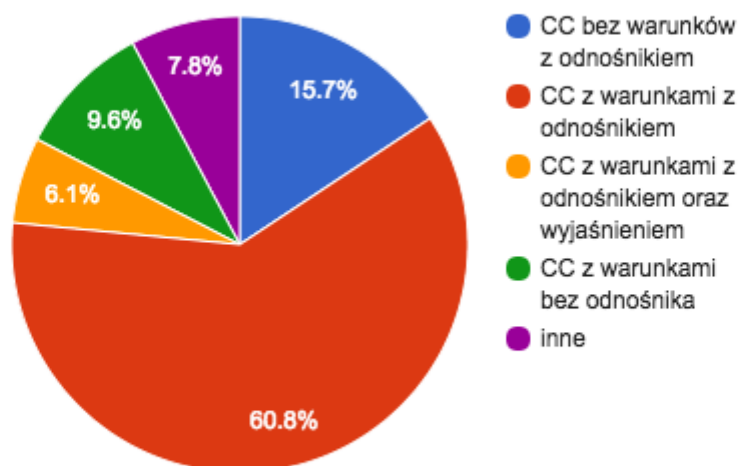


wykres 1.

Przedmiotem naszych zainteresowań był również szczegółowy sposób opisywania licencji CC, na której udostępnia się stronę. Braliśmy pod uwagę, czy nazwana została konkretna licencja - słownie lub za pomocą ikony - oraz czy zamieszczono odnośnik do strony z pełną treścią licencji, na serwerze creativecommons.org. Poprawne opisanie licencji jest wymogiem jedynie dla licencjobiorców, i nie dotyczy podmiotów licencjonujących treści - w tym analizowanych przez nas wypadków licencjonowania treści stron przez ich autorów lub wydawców. Poprawne oznaczanie licencjonowanych treści jest jednak dobrą praktyką rekomendowaną przez CC.

Bardzo różne sposoby opisywania stało się dla nas przyczynkiem do refleksji nad rozumieniem CC przez osoby, które już te licencje wykorzystują i świadomie lub nie dołączają do sieci dzielenia się treściami.

Opis licencji Creative Commons, na której udostępniona jest domena [n=115]



wykres 2.

60% domen oznacza treści licencją poprzez podanie nazwy lub symbolu konkretnej licencji oraz odnośnika do strony z pełną treścią licencji. Spotykaliśmy jednak również sytuacje, w których nie tylko nie wymieniono licencji, ale też linka. To sprawia, że gość takiej strony nie wie, na jakich warunkach może korzystać z udostępnionych tam treści. Z drugiej strony świadczy to o niezrozumieniu systemu/struktury licencji CC przez osoby, które w ten sposób opisują swoje strony.

60% stron poprawnie i kompleksowo wskazujących na wykorzystywaną licencję to dobry wynik, sugerujący wysoką skalę zrozumienia sposobu licencjonowania CC w tej grupie. Jednak aż 20% stron (bez warunków z odnośnikiem (ew. z wyjaśnieniem) (16%) oraz bez warunków i bez odnośnika (4%)) popełnia podstawowy błąd, jakim jest nie wskazanie konkretnej licencji, która została wykorzystana.

Co ciekawe, aż 9% stron prezentuje rozbudowane informacje, wyjaśniające działanie licencji Creative Commons. Oto przykłady takich zapisów:

“Publikujemy za darmo wedle zasady "darmo dostateś (talent) - darmo rozdawaj." Wymaga to udzielenia zgody na publikację na licencji Creative Commons (system zapyta o taką zgodę). W zamian możecie liczyć na setki "pobrań" Waszej twórczości. Powodzenia!

Czytelnicy! Wszystkie publikowane utwory mogą być przez Was swobodnie pobierane do osobistych komputerów zgodnie z zakresem licencji Creative Commons, której udzielają Autorzy”⁷;

“Szablony udostępnione są na licencji Creative Commons - Attribution 2.5. Możesz ich używać w prywatnych jak i komercyjnych projektach. Jedynym wymogiem jest uznanie naszego autorstwa poprzez pozostawienie w widocznym miejscu strony linku do naszego serwisu lub serwisów partnerskich”⁸;

“Wszystkie teksty umieszczone na tej stronie są domyślnie na licencji Creative Commons Uznanie autorstwa - Użycie niekomercyjne - Na tych samych warunkach, co pokrótce znaczy, że każdy fragment tekstu pochodzący z tego miejsca, można używać w każdy sposób, jaki kto chce, tak długo, jak zostanie podany pierwotny autor utworu, nie będzie się z utworu czerpać korzyści majątkowych i każdą przeróbkę tego utworu oddawać społeczeństwu na tej samej licencji, na której go otrzymano”⁹.

Najczęściej wykorzystywaną licencją Creative Commons, w prawie 1/4 przypadków, była BY-SA (uznanie autorstwa - na tych samych warunkach) wykorzystywana również przez Wikipedię.

Niewiele rzadziej wybierano najbardziej restrykcyjne licencje BY-NC-SA (uznanie autorstwa - wykorzystanie niekomercyjne - na tych samych warunkach) i BY-NC-SA (uznanie autorstwa - wykorzystanie niekomercyjne - bez utworów zależnych) oraz równocześnie podstawową, najbardziej liberalną, jedynie z warunkiem uznania autorstwa (BY). Ciekawe, że cztery najpopularniejsze w polskim internecie licencje

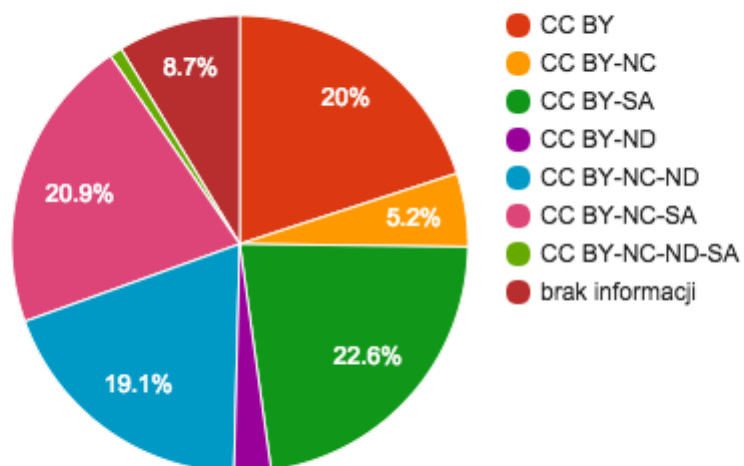
⁷ źródło: <http://www.muvok.com/>, dostęp: 11.08.2015.

⁸ źródło: <http://www.templatki.net/>, dostęp 11.08.2015.

⁹ źródło: <http://pisarczyki.pl/licencje/>, dostęp 11.08.2015.

Creative Commons są najmniej i najbardziej restrykcyjne. Warto też odnotować że prawie 50% stron jest dostępnych na jednej z dwóch wolnych licencji (CC BY lub CC BY-SA).

**Na której licencji udostępniona jest domena?
[n=115]**



wykres 3.

STRONY INTERNETOWE Z TREŚCIAMI UDOSTĘPNIANYMI NA LICENCJACH CREATIVE COMMONS

Przeważającą część treści udostępnianych na licencjach CC stanowią zdjęcia (49%). Kolejne kategorie są udostępniane dużo rzadziej. 12% to Ebooki i pliki tekstowe (zazwyczaj w formacie PDF), a 11% to teksty publikowane bezpośrednio na stronie, np. posty na blogu.



wykres 4.

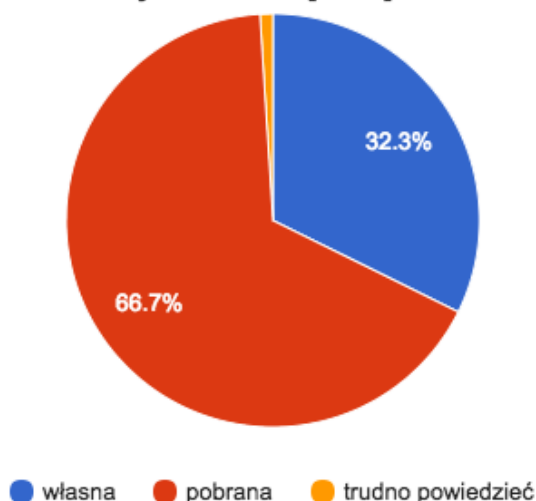
Na powyższym wykresie wyraźnie zatem widać, który rodzaj treści najczęściej jest udostępniany na licencji Creative Commons. Jednym z powodów takiej sytuacji jest rola największych platform, które udostępniają zdjęcia na licencjach CC, które są potem wykorzystywane na innych stronach. Flickr czy Wikimedia Commons dają możliwość lub wymuszają udostępnienie na wybranej (Flickr) lub konkretnej (Wikimedia Commons) licencji CC. Szczególnie Flickr ze swoim przystępnym

/ Przetłaczamy społeczeństwo na cyfrowe

wyjaśnieniem znaczenia poszczególnych licencji (szukanie zdjęć według możliwości wykorzystania komercyjnego i/lub modyfikacji dzieła) nieporównywalnie przyczynia się do rozpowszechniania treści na CC. To zatem bezpośrednio łączy się z wynikami prezentowanymi poniżej. $\frac{2}{3}$ analizowanych treści nie zostało wytworzone przez autorów poszczególnych stron internetowych, a pobrane z innych źródeł. Najczęstszym źródłem był właśnie Flickr. Nawet, gdy pod zdjęciem brakowało informacji o licencji Creative Commons, znajdowała się ona w metadanych zdjęcia i właśnie w ten sposób została zindeksowana przez internetowego robota.

1/3 treści na licencjach Creative Commons została wytworzona przez autorów analizowanych stron WWW.

Czy treść udostępniana na licencji Creative Commons jest treścią wytwarzaną przez autorów strony internetowej, czy jest pobierana z innych źródeł? [n=96]



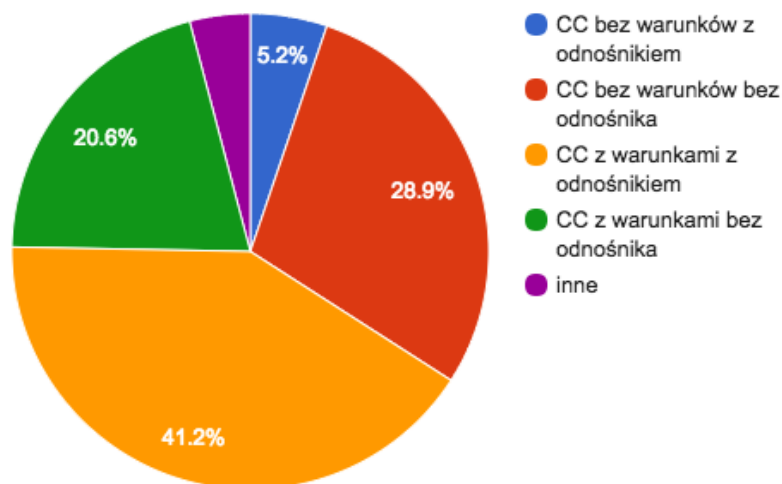
wykres 5.

Podobnie jak przy okazji całych stron internetowych udostępnionych na CC, rozpowszechniane treści posiadają najczęściej zarówno informację o warunkach licencji oraz odnośnik do strony z pełną treścią licencji. Skala jest jednak mniejsza - w przypadku stron internetowych to ponad 60%, a w przypadku treści tylko nieco ponad 40%. Dodatkowo grupa treści, które nie posiadają informacji, na której licencji są udostępniane oraz nie zostały zlinkowane do strony CC, która tę informację by przekazywała, jest zdecydowanie większa - mniej niż 5% przy stronach internetowych oraz prawie 30% przy poszczególnych treściach. Większa jest też grupa treści, dla których podano warunki licencyjne, ale brak odnośnika do pełnej treści licencji. Przy stronach internetowych ta grupa liczyła niecałe 10%, przy treściach jest to dwukrotnie więcej. Widać więc, że licencjonując całe serwisy ich autorzy bardziej

/ Przetłaczamy społeczeństwo na cyfrowe

dbają o poprawność informacji licencyjnych, niż w przypadku licencjonowania pojedynczych treści na stronie. Można zatem dojść do wniosku, że udostępnienie całej strony internetowej na licencji CC jest decyzją bardziej świadomą i zakładającą większą wiedzę o zasadach, na których powinno się to stać.

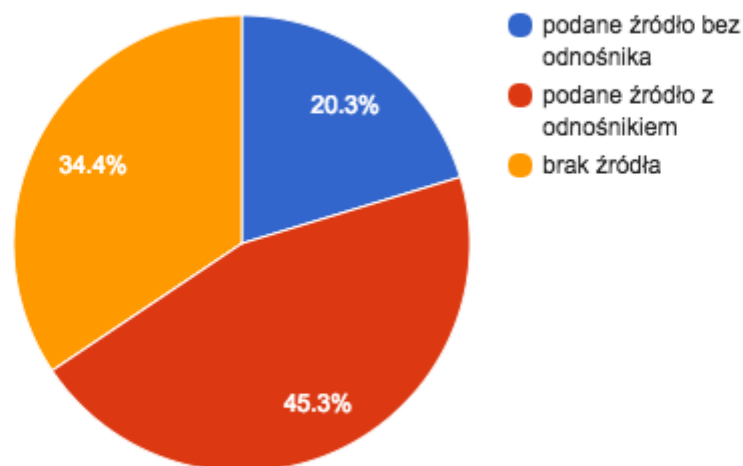
W jaki sposób jest opisana licencja Creative Commons, na której udostępniana jest treść? [n=103]



wykres 6.

Jednym z warunków prawidłowego udostępniania cudzych treści na licencji CC jest podanie źródła. W ponad $\frac{2}{3}$ przypadków tak się stało, choć w co piątej treści brakowało odnośnika do pierwotnie umieszczonej treści. Równocześnie w przypadku $\frac{1}{3}$ treści brakowało jakiegokolwiek informacji o źródle utworu.

Sposób opisu źródła pobranego utworu na licencji Creative Commons [n=64]

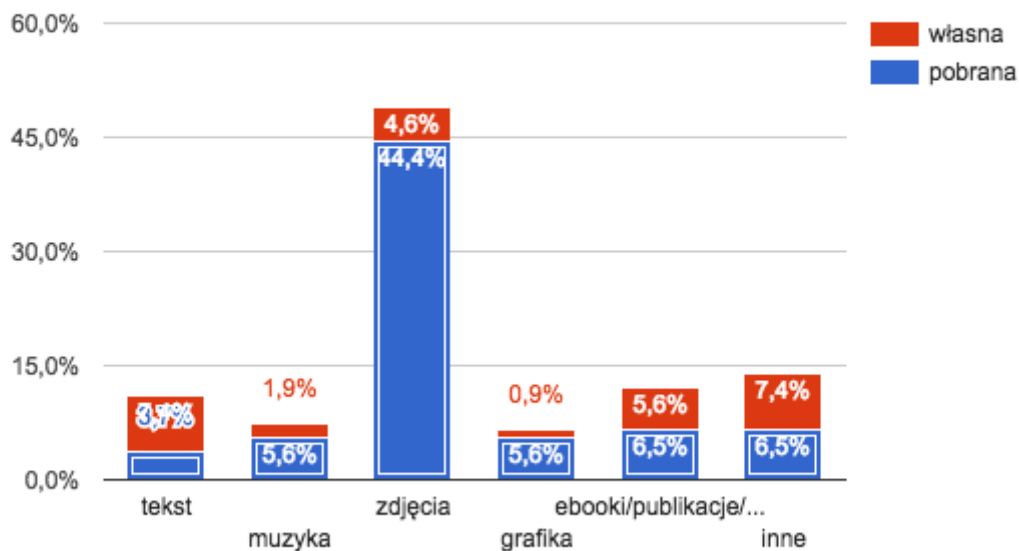


wykres 7.

Znowu dużą rolę w podawaniu źródeł pełniły wspomniane już największe platformy, które automatycznie razem ze skopiowanym zdjęciem umieszczają link do źródła - swojej strony. Treści pobrane to przede wszystkim zdjęcia - pliki ze stron np. Flickr czy Wikimedia Commons.

Potencjalnie porównywalnie dużą bazą plików na CC byłyby platformy filmowe, np. You Tube. Umieszczanie filmów bezpośrednio z takich serwisów nie jest jednak tak rozpowszechnione i łatwe do zrobienia jak w przypadku zdjęć, które w naszym badaniu odgrywały najważniejszą rolę.

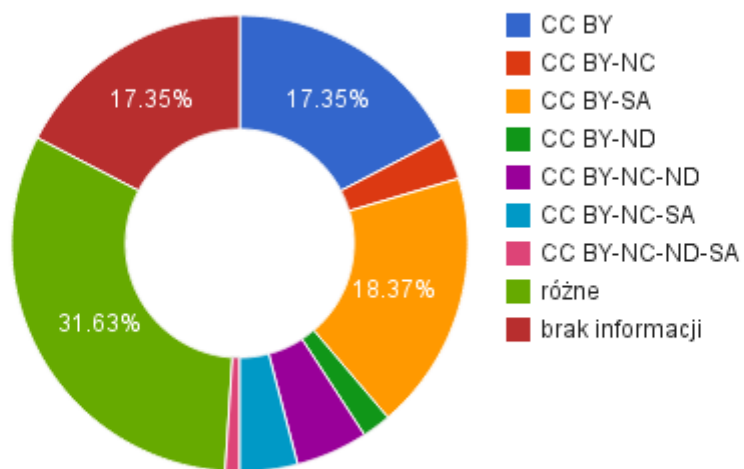
Jaka treść udostępniana jest na licencji Creative Commons? [n=108]



wykres 8.

Ze względu na to, że treści w zdecydowanej większości pochodzą z innych niż dana strona internetowa źródeł, 1/3 stron internetowych zawiera treści udostępnione na różnych licencjach Creative Commons. Gdy zastosowana została jedna licencja, były to CC BY-SA lub CC BY.

na której licencji udostępniane są treści? [n=98]



wykres 9.

SYSTEMATYCZNE CZY INCYDENTALNE WYKORZYSTYWANIE LICENCJI CREATIVE COMMONS

243 544 strony dostępne na licencji CC bądź zawierające treści na tej licencji są nierównomiernie rozłożone pomiędzy domenami. Prezentujemy to na poniższym wykresie. 916 domen posiadało tylko jeden odnośnik związany z Creative Commons, 825 domen posiadało od 2 do 49 linków na stronie. To jeden z ważniejszych wniosków niniejszego badania. Korzystanie z licencji Creative Commons okazuje się być w połowie przypadków zachowaniem incydenalnym, albo nawet jednorazowym. Należy przy tym wziąć pod uwagę, że licencjonowane treści to bardzo często zdjęcia i grafiki, te zaś są w przeważającej mierze pobierane z największych serwisów ze zdjęciami i grafikami (takich jak Flickr oraz Wikimedia Commons). Treści ze wspomnianych portali posiadają informacje o CC w metadanych, a przy kopiowaniu na własną stronę internetową automatycznie zostaje pobrany podpis oraz odnośnik do strony Creative Commons. Mogą być zindeksowane przez roboty internetowe, ale użytkownicy mogą nawet nie wiedzieć, że korzystają z utworów na licencji CC.



wykres 10.

WNIOSKI I REKOMENDACJE

Nasze badanie potwierdziło empirycznie przyjmowaną powszechnie skalę dostępności treści na licencjach Creative Commons - około 1% wszystkich treści. Ten relatywnie mały poziom oznacza ogromną ilość treści dostępnych na wolnych licencjach. Jest również wynikiem, co istotne, świadomej decyzji osób, które decydują się udostępnić treści inaczej niż na działającej automatycznie regule "pełne prawa zastrzeżone". Istotnym wnioskiem jest również to, że połowa treści jest dostępna na jednej z dwóch wolnych licencji: CC Uznanie autorstwa albo CC Uznanie autorstwa - Na tych samych warunkach.

Kluczowe wnioski z analizy dotyczą sposobów udostępniania treści. 60% wszystkich treści jest prawidłowo oznaczonych - zawierają pełną nazwę licencji wraz ze wskazanymi warunkami oraz odnośnik do strony z tekstem licencji. To pozytywny wynik, świadczący o tym, że osoby decydujące się na udostępnianie treści na licencjach CC mają relatywnie wysoki poziom wiedzy prawnoautorskiej. Równocześnie, pozostałe przypadki obejmują jednak sytuacje, w których treści są - na różne sposoby - niepoprawnie oznaczane. Znaczącą rolę może więc odegrać dalsza edukacja i podnoszenie świadomości poprawnego wykorzystywania licencji CC.

W przypadku wykorzystywania cudzych treści, dostępnych na licencjach CC, poziom poprawnych oznaczeń spada do 40%. To istotny wynik, wskazujący, że osoby, które same nie udostępniają treści dużo słabiej rozumieją zasady stosowania tego rodzaju licencji. Wiąże się to z faktem, że wykorzystywanie treści na licencjach CC jest często zachowaniem jednorazowym i incydentalnym.

Ważnym wynikiem są też dane o dominującej roli zdjęć jako treści udostępnianych i wykorzystywanych na licencjach CC. Istotną rolę w przypadku zdjęć odgrywają duże platformy internetowe, takie jak Wikimedia Commons i Flickr. Ich wyraźny wpływ na wykorzystywanie zdjęć na licencjach CC przez internautów wskazuje na dużą rolę, jaką może odgrywać wprowadzanie licencjonowania CC jako funkcjonalności kluczowych serwisów internetowych.

Dalsze badania powinny wyjść poza ilościową analizę stosowania licencji CC, by móc odpowiedzieć na pytania dotyczące motywacji osób udostępniających i wykorzystujących treści, także odnośnie wybieranej licencji.

BIBLIOGRAFIA

Artykuły i notatki

- Creative Commons (2014). State of the Commons: Notes and additional resources.
<https://github.com/creativecommons/stateofthe/blob/master/data/notes.md>
- Creative Commons (2012). The 2012-13 Creative Commons Research agenda.
<https://wiki.creativecommons.org/wiki/Research>
- Creative Commons. Describing Copyright in RDF.
<https://creativecommons.org/ns>
- Tim Hwang (2008). Creative Commons Launches Metrics Research Project.
<https://creativecommons.org/press-releases/entry/8456>

Prezentacje

- Giorgos Cheliotis (2007). Creative Commons Statistics from the CC-Monitor Project. <https://hoikoinoi.files.wordpress.com/2007/07/cc-monitor-findings-isummit.pdf>

Strony internetowe

- <http://pisarczyki.pl/licencje/>
- <http://www.templatki.net/>
- <http://www.muvok.com/>