
Dekonstrukcja artykułu naukowego. Ontologie w publikowaniu semantycznym

Marcin Roszkowski

ORCID 0000-0001-7396-4685

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii,
Uniwersytet Warszawski*

Abstrakt

Cel/Teza: Celem artykułu jest charakterystyka trzech ontologii opracowanych na potrzeby semantycznego publikowania, których przedmiotem opisu jest artykuł naukowy: SciAnnotDoc, Scholarly Papers Vocabulary with Focus on Qualitative Analysis, Document Components Ontology.

Koncepcja/Metody badań: Przeprowadzona charakterystyka ontologii wpisuje się w koncepcję oceny ontologii opartej na interpretacji obecnych w niej postulatów znaczeniowych. Charakterystyka każdej ontologii obejmuje określenie jej zakresu tematycznego, kontekstu powstania, podstawowych założeń ontologicznych oraz próbę ujawnienia jej postaw epistemicznych.

Wyniki i wnioski: Charakterystyka struktur pojęciowych leżących u podstaw trzech ontologii sieciowych, których celem była reprezentacja artykułu naukowego na potrzeby semantycznego publikowania daje obraz modelu konceptualnego tego artefaktu naukowego, w którym przede wszystkim eksponuje się elementy pełniące określoną funkcję retoryczną. W przeanalizowanych przypadkach model IMRaD nie był podstawowym schematem organizacji treści artykułu naukowego. Ujawnienie postaw epistemicznych w procesie projektowania ontologii nie było możliwe we wszystkich przypadkach. Tam, gdzie jednak udało się to określić, widać zarówno postawy obiektywistyczne, jak i interpretatywne, a także obecność determinantów o charakterze pragmatycznym.

Oryginalność/Wartość poznawcza: Modelowanie konceptualne, będące jednym z początkowych etapów projektowania ontologii, jest zdeterminowane określoną postawą epistemiczną, tzn. stosunkiem projektanta do rzeczywistości, który za pomocą ontologii stara się odwzorować jej fragment. Ujawnienie takich postaw jest istotne z punktu widzenia zrozumienia kontekstu postulatów znaczeniowych obecnych w ontologiach sieciowych.

Słowa kluczowe

Artykuł naukowy. Modelowanie konceptualne. Ontologie sieciowe. Postawy epistemiczne. Publikowanie semantyczne.

Otrzymano: 5 czerwca 2019. Zrecenzowano: 29 czerwca 2019. Poprawiono: 4 lipca 2019. Zaakceptowano: 6 lipca 2019.

1. Wprowadzenie

Koncepcja otwartej nauki wprowadziła szereg zmian o charakterze społeczno-kulturowym i technologicznym związanych ze sposobem projektowania, prowadzenia, dokumentowania, rozpowszechniania oraz oceny badań naukowych (Vicente-Saez & Martinez-Fuentes, 2018, 428). Lyubomir Penev (2017, 5) proponuje interpretację otwartej nauki w pięciu kategoriach transformacji, na którą wpływ mają sieciowe i cyfrowe technologie informacyjne. Są to:

- (1) Od otwartego dostępu do otwartej nauki – otwarty dostęp do publikacji i danych badawczych staje się praktyką otwartej nauki i będzie miał wpływ na cały ekosystem produkcji, komunikacji i ponownego użycia osiągnięć naukowych.
- (2) Od treści zrozumiałych dla człowieka do treści automatycznie przetwarzalnych – przystosowanie publikowanych treści naukowych do automatycznego przetwarzania jest obecnie tak samo ważne, jak ich interpretacja przez czytelnika. Pozwala bowiem na ich automatyczne agregowanie oraz zastosowanie technik analizy tekstu.
- (3) Od otwartych danych do ponownego ich wykorzystania – wydawcy powinni wdrażać technologie pozwalające na integrację danych badawczych ze strukturą narracyjną publikacji.
- (4) Od publikowania tradycyjnego do usług informacyjnych opartych na nowoczesnych technologiach – innowacje technologiczne mają kluczowe znaczenie dla właściwego publikowania i rozpowszechniania treści cyfrowych a zatem dla przetrwania i zrównoważonego rozwoju czasopism naukowych i ich wydawców.
- (5) Od semantycznych adnotacji do semantycznego publikowania – wzbogacanie treści publikacji naukowych o formalnie specyfikowane adnotacje jest postrzegane jako krok w kierunku nowej generacji publikowania treści naukowych opartych na technologiach semantycznych.

Obok kwestii związanych ze zmianą paradygmatu obiegu informacji w nauce i roli technologii cyfrowych w tym procesie, mamy tutaj do czynienia ze zmianą podejścia do koncepcji publikacji naukowej. Jest ona traktowana jako złożony artefakt naukowy, dla którego można formalnie zidentyfikować elementy pełniące określoną funkcję retoryczną i prowadzące użytkownika przez proces interpretacji treści, oraz któremu towarzyszą (w uzasadnionych sytuacjach) dane badawcze będące przedmiotem dociekań naukowych. Takie postrzeganie publikacji naukowej nawiązuje do koncepcji publikowania semantycznego (ang. *Semantic Publishing*), która od kilku lat jest intensywnie rozwijana jako kolejny etap ewolucji publikowania i rozpowszechniania treści naukowych w środowisku sieciowym.

Przez semantyczne publikowanie rozumie się reprezentację treści i formy publikacji naukowej za pomocą sieciowych narzędzi organizacji i reprezentacji wiedzy w celu identyfikacji jej elementów strukturalnych oraz warstwy semantycznej. Tego rodzaju wzbogacanie warstwy prezentacji publikacji naukowej o formalnie specyfikowane metadane w ramach platformy wydawniczej lub innej postaci kolekcji cyfrowej zdaniem Davida Shottona (2009, 86) ma pozwolić na automatyczne odkrywanie (ang. *discovery*) publikacji powiązanych semantycznie, dostęp do danych badawczych z poziomu publikacji, czy też integrację danych badawczych pochodzących z wielu publikacji.

Wspomniane wzbogacanie zachodzi poprzez dekonstrukcję publikacji naukowej za pośrednictwem ontologii sieciowych lub formalnie specyfikowanych schematów metadanych oraz poprzez identyfikację i reprezentację pojęć występujących w tekście. Ostatni element jest określany mianem semantycznych adnotacji (ang. *semantic annotation*) i może przebiegać w sposób zautomatyzowany z wykorzystaniem narzędzi eksploracji tekstu (ang. *text-mining*), np. poprzez identyfikację nazw własnych (ang. *named-entity recognition*) lub za pośrednictwem społeczności funkcjonującej w ramach infrastruktury badawczej.

Z formalnego punktu widzenia semantyczna publikacja ma dwie warstwy prezentacji – pierwszą, przeznaczoną dla czytelnika, drugą – zawierającą formalną reprezentację jej treści przeznaczoną do automatycznego przetwarzania. Takie podejście interpretacyjne oddaje

istotę semantycznego publikowania treści naukowych w środowisku sieciowym i nawiązuje do koncepcji World Wide Web. Ta bowiem zakłada rozdzielenie warstw technologicznych odpowiedzialnych za prezentację informacji oraz ich zapis i dystrybucję. Semantyczne publikowanie jest związane z kolejnym etapem ewolucji World Wide Web, jakim jest Sieć Semantyczna (ang. *Semantic Web*), w której przedmiotem opisu nie są dokumenty czy strony internetowe, lecz fakty zawarte w ich treści. U podstaw semantycznego publikowania leży krytyka współczesnego systemu publikowania naukowego w środowisku sieciowym, w którym publikacje, a w szczególności artykuły naukowe, nie zawierają formalnej warstwy semantycznej, która mogłaby zostać przetworzona, zagregowana i zinterpretowana za pośrednictwem technologii cyfrowych (Kuhn & Dumontier, 2017, 139).

Koncepcja semantycznego publikowania opiera się na postrzeganiu publikacji naukowej jako zbioru wzajemnie powiązanych danych odwzorowanych za pomocą technologii i standardów sieciowych. Określa ona również w bezpośredni sposób kluczową rolę środowiska World Wide Web w obiegu informacji w nauce. Semantyczne publikowanie jest również pewnym obszarem badawczym o charakterze implementacyjnym, którego przedmiotem jest identyfikacja i opracowanie narzędzi reprezentacji i organizacji wiedzy właściwych dla obiegu informacji w nauce. Chociaż zakłada się dążenie do formalnej reprezentacji publikacji naukowej jako takiej, to sama inicjatywa oraz zainteresowanie tą formą rozpowszechniania informacji naukowej jest związane z naukami przyrodniczymi i ścisłymi, w których teksty naukowe charakteryzują konstrukcje retoryczne łatwiej poddające się formalizacji niż np. teksty z zakresu nauk społecznych i humanistycznych. Z punktu widzenia form prezentacji treści naukowych w naukach przyrodniczych i ścisłych, to artykuł naukowy (w czasopiśmie czy w materiałach konferencyjnych) jest dominującym typem dokumentu. Stąd też to taka forma publikacji naukowej jest głównym przedmiotem badań w ramach semantycznego publikowania.

Termin *semantyczne publikowanie* w rozumieniu tutaj przedstawionym został upowszechniony przez Davida Shottona, dzisiaj już emerytowanego, chociaż nadal aktywnego, profesora bioinformatyki z Uniwersytetu Oksfordzkiego. Do pierwszych publikacji prezentujących tę koncepcję należy artykuł jego autorstwa z 2009 r. (Shotton, 2009), w którym zaproponował przytoczoną wcześniej definicję oraz nakreślił kontekst pojawienia się potrzeby „semantyzacji” warstwy prezentacji publikacji naukowych w czasopismach online. Jednak pierwsze próby dodania formalnej semantyki do publikacji naukowych są związane z projektem ustrukturyzowanych abstraktów cyfrowych – Structured Digital Abstract (SDA) (Gerstein et al., 2007; Kuhn & Dumontier, 2017, 144). Na poziomie koncepcyjnym projekt ten nawiązuje do abstraktu ustrukturyzowanego, jednak na poziomie wdrożeniowym polega na odwzorowaniu za pomocą specyfikacji XML zarówno jego elementów strukturalnych, jak i pojęć występujących w jego treści. Założeniem twórców SDA było danie autorom możliwości opracowania abstraktu za pomocą odpowiedniego narzędzia w procesie zgłaszania publikacji do wydawnictwa. SDA jest jednak związane z czasopismami z obszaru biologii, czego konsekwencją jest charakterystyczna konstrukcja schematu pojęciowego leżącego u podstaw tej specyfikacji, który jest zdeterminowany dziedzinowo. Za kolejny krok w stronę formalnej reprezentacji treści publikacji naukowej można uznać projekt Journal Article Tag Suite (JATS)¹, który jest rozwijany przez Narodową Bibliotekę Medyczną

¹ <https://jats.nlm.nih.gov/>

w Stanach Zjednoczonych. Jest to również specyfikacja w języku XML, w tym przypadku całej struktury publikacji naukowej. JATS identyfikuje cztery elementy strukturalne tekstu naukowego (ANSI/NISO Z39.96-2019, 2018):

- (1) Front matter – zestaw elementów zawierających metadane na temat publikacji.
- (2) Body matter – zestaw elementów reprezentujących zasadniczą treść publikacji, wewnątrznie ustrukturyzowaną za pomocą specyfikacji JATS (np. paragraf, sekcja, rysunek, tabela).
- (3) Back matter – zestaw elementów służących do reprezentacji aparatu pomocniczego publikacji, np. bibliografii załącznikowej, załączników, indeksu rzeczowego.
- (4) Floating material – zestaw elementów służący do reprezentacji obiektów osadzanych w treści publikacji (np. rysunki, tabele), które decyzją wydawcy mogą być przechowywane niezależnie od treści publikacji.

JATS, chociaż został opracowany na potrzeby nauk medycznych i spopularyzowany przez PubMed Central, cieszy się również zainteresowaniem wydawców spoza tego obszaru, np. Taylor & Francis (Lizzi, 2017). Zarówno SDA, JATS jak i inne inicjatywy (np. DocBook) oparte na odwzorowaniu treści publikacji naukowej za pomocą języka XML, zmierzały do formalnego ujawnienia jej elementów strukturalnych. Kuhn i Dumontier (2017) wskazują na mikropublikację (ang. *micropublication*) i nanopublikację (ang. *nanopublication*) jako kolejne kroki w ewolucji semantycznego publikowania, które tym razem zmierzały w stronę reprezentacji warstwy retoryczno-dyskursywnej tekstu naukowego. Obydwie inicjatywy wyrosły na gruncie nauk medycznych i biologicznych, a ich celem było zaproponowanie ontologii lub formalnych specyfikacji metadanych, które byłyby właściwe do odwzorowania treści publikacji naukowych z tych dziedzin. Dla obecnego stanu rozwoju semantycznego publikowania szczególnie istotne znaczenie ma koncepcja nanopublikacji. W tym przypadku publikację naukową dekonstruuje się do postaci formalnie specyfikowanego zbioru deklaracji reprezentujących tezy zawarte w jej treści. Takie podejście wskazuje na wysoki poziom szczegółowości (ang. *granularity*) reprezentacji wiedzy. Elementarną jednostką opisu nie jest tutaj element strukturalny w postaci sekcji, akapitu czy nawet zdania, lecz fakt naukowy (ang. *claim*), któremu towarzyszy jego uzasadnienie (informacje kontekstowe, wskazanie metody badawczej, narzędzia pomiaru) oraz odwołanie do publikacji, w której treści się pojawił. Do zapisu danych stosuje się tutaj standard reprezentacji wiedzy dla Sieci Semantycznej, którym jest język Resource Description Framework (RDF). Takie podejście do reprezentacji dyskursu naukowego jest obarczone określoną postawą epistemologiczną, która zakłada możliwość jawnej i obiektywnej identyfikacji pojęcia faktu naukowego w tekście publikacji, co w naukach empirycznych jest uzasadnione, ale w innych przypadkach może budzić wątpliwości. Nanopublikacje są rozwijane w naukach biologicznych i medycznych, jednak poza nielicznymi projektami² nie zyskały większego zainteresowania poza nimi. To podejście do reprezentacji publikacji naukowych przyczyniło się jednak do rozwoju koncepcji semantycznego publikowania ze względu na wykorzystanie standardów Sieci Semantycznej (ontologie, język RDF). Środowisko badaczy z obszaru semantycznego publikowania zwróciło uwagę na możliwości, które kryją się w ontologii jako narzędziu reprezentacji wiedzy w kontekście odwzorowania cech formalnych i treściowych publikacji naukowej. Obecnie ruch semantycznego publikowania cechują dwa podejścia. Pierwsze

² <http://nanopub.org/>

polega na wykorzystaniu ontologii i formalnie specyfikowanych schematów metadanych do opisu publikacji naukowej na poziomie strukturalnym i dyskursywno-retorycznym oraz drugie – na wykorzystaniu prostszych narzędzi reprezentacji wiedzy w postaci znaczników semantycznych (ang. *semantic markup*), które stanowią rozszerzenie języka HTML i są zagnieżdżane bezpośrednio w strukturze dokumentu internetowego. W pierwszym przypadku mówimy o konstrukcji baz wiedzy opartych na ontologiach, które pozwalają na przeprowadzanie zaawansowanych analiz i zastosowanie złożonych kryteriów wyszukiwania informacji. W drugim, celem jest przede wszystkim zwiększenie wyszukiwalności publikacji naukowych w środowisku sieciowym poprzez zapewnienie dodatkowych informacji o strukturze i treści publikacji na potrzeby indeksowania przez wyszukiwarki internetowe i agregowania treści przez usługi sieciowe.

O ile ontologie sieciowe należą do najbardziej ekspresywnych sieciowych systemów organizacji wiedzy (Mazzocchi, 2017), to ich konstrukcja jest zadaniem złożonym i czasochłonnym, które wymaga zarówno kompetencji z obszaru reprezentacji wiedzy, jak również szczegółowego poznania dziedziny/domeny, która będzie przedmiotem modelowania konceptualnego. Dodatkowo, ontologia jako sformalizowana teoria pewnej dziedziny wymaga przyjęcia określonej i jawnej postawy poznawczej wobec fragmentu rzeczywistości, który będzie w niej odwzorowany. Mamy więc do czynienia z aspektem epistemologicznym w projektowaniu systemu reprezentacji wiedzy, który wraz z przyjętą metodologią determinuje zarówno proces poznawczy, jak i jego efekt w postaci gotowego narzędzia.

2. Cel i metody badawcze

Głównym celem artykułu jest charakterystyka ontologii opracowanych na potrzeby semantycznego publikowania, których przedmiotem opisu jest artykuł naukowy. Przez artykuł naukowy tutaj rozumiany jest niesamoistny wydawniczo typ dokumentu w postaci artykułu w czasopiśmie, pracy zbiorowej lub w materiałach konferencyjnych. Decyzja o wyborze przedmiotu analiz była podyktowana próbą odpowiedzi na pytanie, w jaki sposób treść i forma jednego z podstawowych artefaktów naukowych w komunikacji naukowej jest odwzorowywana za pomocą ontologii sieciowych. Kontekstem do podjętych analiz jest problem ograniczeń schematów metadanych i standardów opracowania dokumentów naukowych wypracowanych w tradycji biblioteczno-bibliograficznej oraz zakres wykorzystania modelu IMRaD (Introduction, Methods, Results and Discussion; wstęp, metodologia, wyniki, dyskusja) jako pewnego rodzaju konsensusu społeczności akademickiej, dotyczącego standardowej struktury artykułu naukowego.

Ontologia w rozumieniu systemu organizacji wiedzy jest definiowana jako formalna i jawna specyfikacja współdzielonej konceptualizacji pewnej domeny (Studer, Benjamins, & Fensel, 1998, 25). Przez konceptualizację będziemy tutaj rozumieli zestaw pojęć zidentyfikowanych w celu wyrażenia abstrakcji na temat stanu rzeczy w określonej dziedzinie lub fragmencie opisywanej rzeczywistości – uniwersum dyskursu (Guizzardi, 2005, 2). Do wyrażenia określonej konceptualizacji potrzebny jest język, który zapewnia zarówno właściwą semantykę, jak i syntaktyczne środki ekspresji znaczenia. Dotyczy to zarówno identyfikacji pojęć i ich zależności hierarchicznych, jak również relacji semantycznych, jakie tworzą. Jawność i formalność specyfikacji w ontologiach zapewnia język reprezentacji

wiedzy, którym dla ontologii sieciowych jest Web Ontology Language (OWL). Specyfikacja danej konceptualizacji z wykorzystaniem języka reprezentacji wiedzy zachodzi na poziomie ekstensjonalnym i intensjonalnym. Z punktu widzenia efektywności odwzorowania pewnej struktury pojęciowej, to właśnie poziom intensjonalny jest właściwy do określania warunków interpretacji zależności między pojęciami. Przyjmuje się zatem, że ontologia jest systemem aksjomatów, tj. logiczną teorią opracowaną na potrzeby odwzorowania modelu pojęciowego w danym projekcie, który reprezentuje zakładaną konceptualizację i w warstwie jej formalnologicznej konstrukcji wyklucza inne możliwe interpretacje (Guarino et al., 2009, 8). O ile formalny charakter specyfikacji jest zapewniony przez język reprezentacji wiedzy, to na poziomie semantycznym proces tworzenia ontologii zakłada konieczność identyfikacji klas reprezentujących pojęcia, atrybutów odwzorowujących ich własności (ang. *data properties*) oraz relacji zachodzących między klasami (ang. *object properties*). Zarówno klasy, atrybuty, jak i relacje mają swoje wykładniki językowe oraz definicje (adnotacje) charakteryzujące ich znaczenie, co zapewnia właściwą interpretację przez użytkownika. Mamy zatem do czynienia z reprezentacją pewnej konceptualizacji na dwóch poziomach, pierwsza – na poziomie języka naturalnego, druga – języka formalnego. Ontologia jest więc przykładem systemu organizacji wiedzy operującego notacją paranaturalną z dodatkową warstwą formalnologiczną, która zapewnia precyzyjne ramy interpretacyjne. Komentarza wymaga również ostatni warunek definicyjny ontologii – współdzielony charakter konceptualizacji. Oznacza to włączenie społecznego determinantu o charakterze interpretacyjnym do procesu tworzenia i wykorzystania ontologii. Chociaż możemy przyjąć, że warunek współdzielenia kompletnej konceptualizacji jest nie do przyjęcia (Guarino et al., 2009, 14), to ontologia powinna być wyrazem konsensusu poznawczego danej społeczności, która będzie z niej korzystać na mocy rozwiązania o charakterze przybliżonym (ang. *approximation*). Guarino i inni wskazują również na możliwość współdzielenia pewnych właściwych dla danej domeny postulatów znaczeniowych (ang. *meaning postulates*), które w ontologiach przyjmują postać formalnych ograniczeń definicyjnych dla pojęć i relacji. Mamy tutaj do czynienia również z aspektem pragmatycznym. Jeżeli ontologia nie spełnia wymogu współdzielenia konceptualizacji, to istnieje prawdopodobieństwo, że nie będzie zrozumiała przez społeczność, do której jest skierowana i jej użyteczność będzie kwestionowana.

Metodologia projektowania ontologii jest dobrze rozpoznana i udokumentowana w piśmiennictwie naukowym (Fernández-López & Gómez-Pérez, 2002; McGuinness & Noy, 2001; Pérez et al., 2008). Problemem jednak może być tutaj wstępna faza projektowania, na którą składa się określenie zakresu tematycznego ontologii oraz identyfikacja kluczowych pojęć i relacji zachodzących w danej konceptualizacji. Na tym etapie wszystkie podjęte decyzje dotyczące tego, „co i jak istnieje” w danej dziedzinie mają konsekwencje w postaci specyfikacji danej ontologii. Mamy więc tutaj do czynienia z epistemologicznym aspektem projektowania ontologii, tzn. z określonym stosunkiem projektanta do rzeczywistości, który za pomocą ontologii stara się odwzorować jej fragment. Problem analizy postaw epistemicznych (ang. *epistemic stance*) w projektowaniu ontologii nie jest dobrze rozpoznany w piśmiennictwie z tego obszaru, chociaż zdaniem autora ma on zasadniczy wpływ na to jak będzie wyglądał ostateczny efekt samej konceptualizacji. Zagadnienie to było już omawiane w piśmiennictwie z obszaru organizacji wiedzy (np. Hjørland, 2003; Zins, 2004) i modelowania konceptualnego systemów informacyjnych (np. Recker & Niehaves, 2008; Ribbert et al., 2004) oraz modelowania danych (Klein & Hirschheim, 1987). W drugim

przypadku przyjmuje się tezę, że model konceptualny jako pewnego rodzaju teoria dziedziny jest zdeterminowany epistemologicznie (Ribbert et al., 2004, 4232). Oznacza to, że sposób postrzegania rzeczywistości oraz metody formułowania wniosków na jej temat są uwarunkowane określoną postawą epistemiczną projektanta lub grupy projektowej. W odniesieniu do modelowania konceptualnego wyróżnia się dwie główne postawy ontologiczne – realizm krytyczny oraz konstruktywizm ontologiczny (Mitterer, 2017, XV). W dużym uproszczeniu, pierwsza zakłada odwzorowanie zróżnicowania rzeczywistości w jej modelu konceptualnym, druga, że wiedza o rzeczywistości jest konstruowana przez podmiot poznający. W badaniach nad modelami konceptualnymi systemów informacyjnych Jörg Becker i Björn Niehaves proponują schemat pięciu pytań epistemologicznych, których odpowiedzi pozwolą ujawnić postawy epistemiczne przyjęte podczas procesu projektowania modelu konceptualnego systemu informacyjnego (Tab. 1).

Schemat Beckera i Niehavesa pozwala zatem określić determinanty poznawcze obecne w procesie modelowania konceptualnego a z uwagi na fakt, że ten proces zachodzi również podczas projektowania ontologii sieciowych, daje to możliwość zastosowania go w badaniach nad tego rodzaju systemami organizacji wiedzy. H. Klein i R. Hirschheim (1987, 9) twierdzą jednak, że postawy epistemiczne w modelowaniu konceptualnym wynikają z dwóch przeciwstawnych paradygmatów – obiektywistycznego i subiektywistycznego i lokują się między podejściem realistyczno-pozytywistycznym a nominalistyczno-interpretatywnym.

Na płaszczyźnie metodologicznej przedstawione w dalszej części rozważania na temat sposobów dekonstrukcji artykułu naukowego w ontologiach sieciowych są oparte na próbie ujawnienia postaw epistemicznych obecnych w tych narzędziach oraz na analizie konsekwencji takich założeń w postaci konstrukcji określonych schematów organizacji wiedzy. W niniejszym artykule nie dążono do przedstawienia kompletnego przeglądu istniejących ontologii, lecz skupiono się na analizie trzech przypadków ontologii, w których artykuł naukowy jest głównym przedmiotem opisu. Badania nad zidentyfikowanymi ontologiami wpisują się w ujęcie interpretatywne oparte na analizie piśmiennictwa na temat poszczególnych ontologii, ich dokumentacji oraz bezpośrednio samych schematów. Taki proces badawczy do pewnego stopnia nawiązuje do metody ewaluacji ontologii Asuncion Gomez-Perez (2013, 252), która zidentyfikowała trzy poziomy ich oceny:

- (1) Weryfikacja (ang. *verification*) – formalna poprawność konstrukcji ontologii.
- (2) Walidacja (ang. *validation*) – semantyczna zbieżność ontologii z modelem konceptualnym dziedziny.
- (3) Ocena (ang. *assessment*) – interpretacja, użyteczność, jakościowy charakter formalizacji konceptualizacji.

W tak przedstawionej koncepcji ewaluacji ontologii, rozważania w tym artykule lokują się w obszarze oceny ontologii w ujęciu interpretatywnym z odniesieniem do założeń epistemologicznych.

Tab. 1. Pytania epistemologiczne w modelowaniu konceptualnym.

Źródło: Niehaves & Becker, 2006, 202³

| | | | |
|---|--|---|---|
| 1. Co jest przedmiotem poznania? (Aspekt ontologiczny) | <i>Realizm ontologiczny.</i> Świat istnieje niezależnie od ludzkiego poznania, np. procesów myślowych i aktów mowy. | <i>Idealizm ontologiczny.</i> Świat jest konstruktem zależnym od ludzkiej świadomości. | <i>Kantianizm.</i> Istnieją elementy zarówno zależne (fenomen), jak i niezależne (noumen) od ludzkiego umysłu. |
| 2. Jaka jest relacja między procesem poznania a przedmiotem poznania? | <i>Realizm epistemologiczny.</i> Obiektywne poznanie niezależnej rzeczywistości jest możliwe. | <i>Konstruktywizm.</i> Relacja między procesem poznania a przedmiotem poznania jest zdeteminowana przez podmiot poznający. | |
| 3. Czym jest prawdziwe poznanie? (Koncepcja prawdy) | <i>Korespondencyjna teoria prawdy.</i> Twierdzenie prawdziwe to takie, które jest zgodne z faktami w rzeczywistości, czyli odpowiada rzeczywistości stanowi rzeczy. | <i>Konsensualna koncepcja prawdy.</i> Twierdzenie jest prawdziwe (dla grupy) jeżeli akceptowane przez grupę. | <i>Semantyczna teoria prawdy.</i> Warunkiem prawdziwości jest rozdzielenie przedmiotu poznania od metajęzyka ⁴ . |
| 4. Jakie są źródła poznania? | <i>Empyryzm.</i> Poznanie zachodzi za pośrednictwem zmysłów. Wiedza oparta na doświadczeniu nazywana jest wiedzą a posteriori lub empyryczną. | <i>Racjonalizm.</i> Poznanie zachodzi za pośrednictwem intelektu. Taki rodzaj wiedzy nazywa się wiedzą a priori. | <i>Kantianizm.</i> Zarówno doświadczenie jak i intelekt są źródłami poznania. |
| 5. W jaki sposób można przeprowadzić proces poznawczy? (Aspekt metodologiczny) | <i>Indukcjonizm.</i> Indukcja jest rozumiana jako uogólnianie, generalizowanie na podstawie indywidualnych przypadków. | <i>Dedukcjonizm.</i> Dedukcja polega na wyprowadzaniu wniosków szczegółowych na podstawie wiedzy ogólnej. | <i>Hermeneutyka.</i> Rozumienie danego zjawiska jest uwarunkowane posiadanym przed-rozumieniem całości lub kontekstu, w jakim funkcjonuje. |

³ Tabela w przekładzie własnym autora.⁴ Widoczne jest tutaj odwołanie do semantycznej teorii prawdy Tarskiego, w której prawda jest funkcją zdania.

3. Rezultaty

Do identyfikacji ontologii sieciowych wykorzystano rejestr ontologii i formalnie specyfikowanych schematów metadanych w postaci bazy Linked Open Vocabulary (LOV)⁵ oraz przeprowadzono analizę piśmiennictwa. W toku przeprowadzonych poszukiwań do szczegółowej analizy wybrano trzy ontologie, których bezpośrednim celem jest reprezentacja artykułu naukowego: The SciAnnotDoc model, Document Components Ontology, Scholarly Papers Vocabulary with Focus on Qualitative Analysis

Analizę każdej ontologii przeprowadzono za pomocą schematu interpretacyjnego, który zakłada określenie zakresu tematycznego ontologii, kontekstu jej powstania, założeń metodologicznych oraz analizę schematu pojęciowego stanowiącego jej fundament. Na podstawie analizy dokumentacji i piśmiennictwa, ze szczególnym uwzględnieniem zastosowanych rozwiązań metodologicznych, starano się wskazać postawy epistemiczne przyjęte przez twórców podczas procesu projektowania.

3.1. *SciAnnotDoc*

Ontologia SciAnnotDoc została opracowana na potrzeby systemu informacyjnego, którego główną funkcjonalnością miała być automatyczna ekstrakcja pojęć z korpusu artykułów naukowych z obszaru gender studies. Ontologia oraz prototyp systemu powstały w wyniku prac grupy badawczej afiliowanej przy Uniwersytecie Genewskim oraz Uniwersytecie Walijskim w Cardiff. Głównym założeniem podczas modelowania konceptualnego, które miało doprowadzić do opracowania ontologii, była dekompozycja artykułu naukowego do postaci elementów strukturalnych i dyskursywnych, w obrębie których miał zachodzić proces identyfikacji pojęć związanych z obszarem gender studies (de Ribaupierre, 2014; de Ribaupierre & Falquet, 2018). Ontologia SciAnnotDoc została opracowana z wykorzystaniem języka OWL i chociaż nie jest opublikowana w repozytorium LOV, to sama jej specyfikacja jest ogólnie dostępna⁶. Proces konceptualizacji poprzedzający konstrukcję ontologii polegał na wykorzystaniu wyników uzyskanych z:

- (a) badania ankietowego przeprowadzonego wśród potencjalnych użytkowników systemu (n=91) dotyczących ich zachowań informacyjnych i oczekiwanych funkcjonalności wobec systemu informacyjnego;
- (b) wywiadów z naukowcami z różnych dyscyplin (n=10);
- (c) automatycznej ekstrakcji danych z abstraktów artykułów z bazy PubMed (n=1500) oraz z czasopism z obszaru gender studies (n=1500).

Na podstawie wniosków z badań, twórcy projektu SciAnnotDoc zdecydowali się na odrzucenie modelu IMRaD jako schematu strukturyzacji artykułu naukowego i przyjęcie własnej konceptualizacji, która zakłada cztery poziomy reprezentacji (de Ribaupierre & Falquet, 2018, 276–277):

- (1) **Metadane** (ang. *metadata*) – zestaw atrybutów artykułu naukowego postrzeganego jako dokument (np. tytuł, autor, tytuł czasopisma);

⁵ <https://lov.linkeddata.es>

⁶ http://users.cs.cf.ac.uk/DeRibaupierreH/sciAnnotDoc_article_data_files/

- (2) **Zawartość tekstowa** (ang. *textual content*) – identyfikacja trzech kategorii pojęć występujących w treści artykułu: *metody* (ang. *methods*; np. nazwy metod, typy zmiennych, narzędzia pomiaru), *obiekty naukowe* (ang. *scientific objects*; np. modele, algorytmy, teorie, zasady), *pojęcia dziedzinowe* (ang. *domain concepts*). SciAnnotDoc stosuje tutaj podejście modułowe, tzn. dla każdej z trzech kategorii pojęć tworzy się osobną ontologię lub sformalizowany system organizacji wiedzy (np. listę wartości, taksonomię, klasyfikację). Daje to pewną elastyczność tej ontologii w kontekście wykorzystania w innym obszarze niż gender studies poprzez wdrożenie zewnętrznego systemu organizacji wiedzy w miejsce modułu pojęć dziedzinowych;
- (3) **Elementy dyskursywne** (ang. *discourse elements*) – podstawowy element ontologii SciAnnotDoc, który tworzą: *wyniki*, *definicja*, *metodologia*, *hipoteza* oraz *stan badań*⁷. Elementy te utworzono na podstawie wyników badań z użytkownikami. Przyjęto również, że na poziomie dyskursywnym artykuł jest zbudowany z *fragmentów strukturalnych* (ang. *structural fragments*; najczęściej paragrafów), które z kolei składają się z *elementów strukturalnych* (jedno lub kilka zdań). Zarówno fragment, jak i element może pełnić określoną funkcję dyskursywną;
- (4) **Elementy relacyjne** (ang. *relational elements*) – system wykładników relacji, które zachodzą pomiędzy dokumentem lub jednym z jego elementów dyskursywnych a innym dokumentem lub jego elementem dyskursywnym. SciAnnotDoc wykorzystuje zewnętrzną ontologię Citation Typing Ontology⁸, w celu odwzorowania różnych typów cytowań, które mogą zachodzić między dokumentami naukowymi. SciAnnotDoc pozwala również na odwzorowanie relacji mereologicznej zachodzącej między *elementem* oraz *fragmentem strukturalnym*.

Rysunek 1 przedstawia wizualizację ontologii SciAnnotDoc w postaci schematu i uwzględnia kluczowe poziomy reprezentacji.

Twórcy SciAnnotDoc odrzucili model IMRaD, uzasadniając to jego małą użytecznością w procesie adnotacji, twierdząc dodatkowo, że wyróżnione w nim sekcje pełnią raczej funkcję retoryczną lub dyskursywną. Dodatkowo na podstawie analizy próby artykułów z czasopism z tego obszaru (n=42) stwierdzili, że nie jest on powszechnie stosowanym narzędziem strukturyzacji treści publikacji (de Ribaupierre, 2014, 74). Ontologia SciAnnotDoc wprowadza jednak pojęcie (klasę) *logicznej struktury* artykułu, która do pewnego stopnia jest z nim zbieżna. Tworzy je 10 elementów:

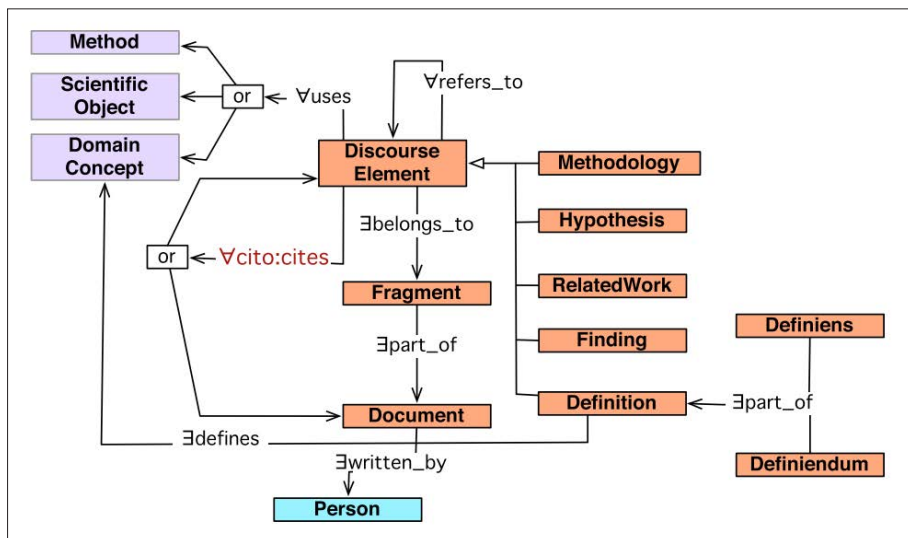
- (1) Abstrakt (ang. *abstract*);
- (2) Tło/kontekst (ang. *background*);
- (3) Wnioski (ang. *conclusion*);
- (4) Dyskusja (ang. *discussion*);
- (5) Ocena (ang. *evaluation*);
- (6) Dalsze badania (ang. *future work*⁹);
- (7) Wprowadzenie (ang. *introduction*);
- (8) Metodologia (ang. *methodology*);

⁷ W oryginale występuje termin *related work*, ale zdaniem autora w polskim piśmiennictwie bardziej rozpowszechnione jest sformułowanie *stan badań*.

⁸ <http://www.sparantologies.net/ontologies/cito>

⁹ Pisownia oryginalna.

- (9) Stan badań (ang. *related work*);
- (10) Rezultaty (ang. *results*).



Rys. 1. Graficzna reprezentacja ontologii SciAnnotDoc.
 Źródło: de Ribaupierre & Falquet, 2018, 277

Mamy więc do czynienia z dekonstrukcją artykułu naukowego na poziomie strukturalnym (struktura logiczna), w ramach której identyfikuje się fragmenty (akapity, ang. *paragraphs*) i elementy strukturalne (zdanie/zdania), które pełnią określoną funkcję dyskursywną. Dodatkowo, element struktury logicznej może zawierać terminy reprezentujące trzy kategorie pojęciowe: metodę, obiekt naukowy, pojęcie dziedzinowe.

W przypadku klasy *struktura logiczna* artykułu naukowego pojawia się pewien problem interpretacyjny. Każdy z dziesięciu elementów przedstawionych powyżej z formalnego punktu widzenia pełni funkcję instancji tej klasy. Oznacza to, że elementy te nie pełnią funkcji generalizacji i nie mają charakteru abstrakcyjnego, lecz stanowią nazwy dla obiektów rzeczywistych. Skoro mają one status instancji, to *de facto* nie można odwzorować tutaj logicznej struktury artykułu poprzez reprezentację kolejności ich występowania w treści. Pełnią one tylko funkcję wyliczającą. Zarówno dokumentacja ontologii, jak i publikacje na jej temat nie zawierają uzasadnienia tej decyzji. Kolejny problem interpretacyjny stanowi zbieżność zakresowa nazw instancji z niektórymi klasami elementów dyskursywnych. Mowa tutaj o parach przedstawionych w tabeli 2.

Tab. 2. Podobieństwo znaczeniowe elementów w SciAnnotDoc

| Elementy dyskursywne | Struktura logiczna |
|--|--|
| Metody (ang. <i>Methods</i>) | Metodologia (ang. <i>Methodology</i>) |
| Wyniki (ang. <i>Findings</i>) | Rezultaty (ang. <i>Results</i>) |
| Stan badań (ang. <i>Related Work</i>) | Stan badań (ang. <i>Related Work</i>) |

Tego typu kwestie dotyczą co prawda formalnej oceny ontologii, jednak z punktu widzenia interpretacji przez użytkownika i potencjalnej użyteczności mają istotne znaczenie. Do tej grupy uwag należą również definicje klas i relacji. Nie wszystkie zidentyfikowane klasy są wyposażone w adnotacje w języku naturalnym. Tylko na podstawie formy wykładnika językowego, jako nazwy klasy oraz miejsca w strukturze paradygmatycznej, użytkownik jest w stanie zinterpretować znaczenie wprowadzonego w tym modelu pojęcia. To samo dotyczy relacji między klasami.

Na podstawie powyższych analiz można pokusić się o próbę określenia postaw epistemicznych przyjętych przez projektantów tej ontologii. Po pierwsze ontologia SciAnnotDoc powstała na potrzeby systemu informacyjnego o określonych wymaganiach funkcjonalnych, zatem aspekt pragmatyczny był tutaj z pewnością obecny przy podejmowaniu wiążących decyzji. Budowa scenariusza badań użytkowników (de Ribaupierre, 2014, 199) pozwala sądzić, że wykorzystano podejście interpretatywne, oparte na koncepcji konsensu. Można przypuszczać, że przedmiotem poznania były tutaj pewne modele mentalne uniwersum dyskursu. Sam proces konstrukcji konceptualizacji miał charakter racjonalistyczny, zaś weryfikacja jego prawdziwości/użyteczności była oparta na podejściu empirycznym. Proces ten zmierzał do zrozumienia, czym kierują się użytkownicy podczas wyszukiwania informacji i jakie elementy strukturalne i dyskursywne są dla nich istotne podczas czytania tekstu naukowego (de Ribaupierre & Falquet, 2018, 276). Chociaż model zakłada pięć poziomów interpretacji, to głównym przedmiotem poznania nie jest tutaj artykuł naukowy w rozumieniu fizycznie utrwalonego artefaktu naukowego, lecz jego warstwa dyskursywna. Analiza klas, atrybutów i pojęć związanych z postrzeganiem artykułu jako jednostki bibliograficznej pokazuje, że mamy tutaj do czynienia z dużym poziomem ogólności. To co jest eksponowane, to elementy dyskursywne oraz pojęcia występujące w ich kontekście.

3.2. Scholarly Papers Vocabulary with Focus on Qualitative Analysis

Ontologia Scholarly Papers Vocabulary with Focus on Qualitative Analysis (SPVQA) została opracowana w ramach prac grupy badawczej BMaKE – Research Group Business Modeling and Knowledge Engineering afiliowanej przy Technische Hochschule Brandenburg w Niemczech. Głównym założeniem projektu było opracowanie systemu informacyjnego na potrzeby infrastruktury badawczej BMaKE, który ułatwiłby analizę zawartości artykułów naukowych (przede wszystkim z materiałów konferencyjnych) z obszaru inżynierii systemów informacyjnych (Meister, 2017, 71) za pośrednictwem procesu manualnego indeksowania (adnotacje) przez użytkowników. Tak jak w przypadku SciAnnotDoc, mamy więc i tutaj do czynienia z determinantem pragmatycznym, dodatkowo wzmocnionym wymaganiami funkcjonalnymi oprogramowania, w ramach którego miała funkcjonować opracowana ontologia. W projekcie zdecydowano się na wdrożenie aplikacji OntoWiki¹⁰, która jest społecznie rozwijaną platformą typu Wiki z możliwością implementacji bazy wiedzy. Ontologia została opracowana w języku RDF i jest zarejestrowana w LOV¹¹.

¹⁰ <http://ontowiki.net/>

¹¹ <https://lov.linkeddata.es/dataset/lov/vocabs/spvqa>

Linked Data. Za podstawową specyfikację służącą do reprezentacji konceptualizacji SPVQA wybrano standard schema.org, który jest obecnie jednym z najczęściej stosowanych semantycznych języków znacznikowych¹². Schemat ten okazał się być adekwatny, ale tylko do reprezentacji podstawowych pojęć w modelu i to bezpośrednio związanych z wymiarem bibliograficznym. Kolejnym krokiem była analiza ontologii Discourse Elements Ontology (DEO, zob. sekcja 3.3.). Pomimo podobnego zakresu tematycznego, zdecydowano jednak nie wykorzystywać elementów tej ontologii z powodu rozbieżności interpretacji elementów retorycznych oraz z powodów formalnych (Meister, 2017, 75) wynikających z faktu, że DEO jest specyfikacją zapisaną w języku OWL a SPVQA wykorzystuje RDF.

SPVQA jest niewątpliwie łatwiejsza do interpretacji dla użytkowników niż SciAnnotDoc. Każda klasa i relacja zidentyfikowana w modelu została opatrzona właściwą definicją lub adnotacją w języku naturalnym. Do nielicznych problemów związanych z interpretacją należy sytuacja reprezentacji problematyki metod badawczych. Ontologia SPVQA¹³ wprowadza cztery pojęcia o różnych statusach służące do odwzorowania zagadnień metodologicznych poruszanych w treści artykułu naukowego (Tab. 3).

Tab. 3. Problematyka metod badawczych w ontologii SPVQA

| IRI | Nazwa | Status | Definicja |
|------------------------|----------------------|---------|---|
| spv#ResearchMethodType | Research Method Type | Klasa | Typy metod badawczych wykorzystywane w badaniach z obszaru inżynierii systemów informacyjnych ¹⁴ |
| spv#ResearchMethods | Research Methods | Klasa | Klasa abstrakcyjna wykorzystywana do modelowania metod badawczych zastosowanych w artykule naukowym ¹⁵ |
| spv#ResearchMethod | Research Method | Klasa | Klasa abstrakcyjna wykorzystywana do modelowania metod badawczych zastosowanych w artykule naukowym ¹⁶ |
| spv#researchMethod | research method | Relacja | Służy do wskazania metody badawczej zastosowanej w artykule naukowym ¹⁷ Dziedzina relacji: Metoda badawcza (ang. <i>Research Method</i>) Zakres wartości: Typy metod badawczych (ang. <i>Research Method Type</i>) |

Na podstawie informacji przedstawionych w tabeli 3 można postawić cztery wnioski:

- (1) Klasa *Metody badawcze* (ang. *Research Methods*) jest zbędna;
- (2) Klasa *Typy metod badawczych* (ang. *Research Method Type*) służy do grupowania nie tyle typów, co konkretnych metod badawczych właściwych dla obszaru inżynierii systemów informacyjnych;
- (3) Klasa *Metoda badawcza* (ang. *Research Method*) pełni funkcję retoryczną;

¹² <http://webdatacommons.org/structureddata/>

¹³ <https://bmake.th-brandenburg.de/spv.rdf>

¹⁴ Oryg. Types of research methods used in Information Systems.

¹⁵ Oryg. Abstract class for modeling a list of research methods applied in an analyzed scientific paper.

¹⁶ Oryg. Abstract class for modeling a list of research methods applied in an analyzed scientific paper.

¹⁷ Oryg. Is used to indicate a research method of a scientific paper.

- (4) Relacja *metoda badawcza* (ang. *research method*) zachodzi między elementem retorycznym artykułu naukowego a konkretną metodą badawczą właściwą dla obszaru inżynierii systemów informacyjnych.

Brak identyfikacji elementów strukturalnych artykułu naukowego w ontologii SPVQA powoduje trudności w interpretacji, co jest elementarną jednostką, o której można formalnie orzec, że dotyczy np. metody badawczej czy celu badawczego. Na podstawie ontologii oraz jej dokumentacji nie można również określić, czy opis będzie realizowany na poziomie całego dokumentu, czy też pojedynczego faktu.

Z punktu widzenia analizy postaw epistemicznych, obecnych podczas projektowania SPVQA, trudno jednoznacznie wskazać na konkretne perspektywy poznawcze. Zarówno dokumentacja, jak i piśmiennictwo na temat tego projektu, nie dostarczają przesłanek do formułowania takich wniosków. Vera Meister (2017, 73) – autorka ontologii formułuje tezę, że głównym celem bazy wiedzy opracowanej za pomocą ontologii SPVQA jest umożliwienie grupie badawczej kolektywnej analizy publikacji z obszaru inżynierii systemów informacyjnych, co czyni **oczywistą** konieczność strukturyzacji artykułów naukowych uwzględniając kryteria jakościowe: (1) cele badawcze, (2) metody badawcze, (3) rezultaty, (4) dalsze badania oraz (5) krytykę¹⁸. Wspomniane tutaj kryteria znalazły swoje odzwierciedlenie w głównej osi koncepcyjnej ontologii SPVQA. Jednak w dalszej części publikacji nie ma wzmianki dotyczącej decyzji metodologicznych związanych z procesem konceptualizacji, które rzuciłyby światło na przyjęte postawy epistemiczne. Nie można zatem określić źródeł wiedzy na temat modelowanej dziedziny. Z przytoczonego fragmentu można jedynie wywnioskować, że przyjęto określone założenia o charakterze apriorycznym, co może świadczyć o podejściu racjonalistycznym, ale równie dobrze może świadczyć o naiwnym realizmie. Dalsze próby określenia determinantów epistemologicznych dla tej ontologii przy braku dodatkowych informacji są skazane wyłącznie na spekulacje.

3.3. Document Components Ontology

Ontologia Document Components Ontology (DOCO) jest jedną z ontologii opracowanych w ramach projektu Semantic Publishing and Referencing Ontologies (SPAR)¹⁹, którego celem jest rozwój i popularyzacja koncepcji semantycznego publikowania poprzez zapewnienie gotowego zestawu narzędzi reprezentacji wiedzy obejmujących cały cykl życia publikacji naukowej i możliwych do zastosowania niezależnie od dziedziny. Sam projekt jest rozwijany od 10 lat przez grupę badawczą kierowaną przez wspomnianego wcześniej Davida Shottona. Głównym założeniem projektu SPAR było opracowanie ekosystemu ontologii o charakterze modułowym, które mogłyby zostać wdrożone całościowo lub indywidualnie, zgodnie z wymaganiami funkcjonalnymi określonego systemu informacyjnego. W odróżnieniu od SciAnnotDoc i SPVQA, założeniem projektu SPAR było opracowanie ontologii o uniwersalnym zakresie stosowania i niezależnych od technicznych wymogów

¹⁸ Oryg: Like stated above, the main objective of the required knowledge base is to support the research group's collective analysis of scientific publications in the field of information systems. It is therefore obvious to structure scholarly papers according to their main qualitative features: (i) research objectives, (ii) research methods, (iii) research findings, (iv) future work, and (v) critical issues.

¹⁹ <http://www.sparontologies.net/>

systemów informacyjnych, w ramach których mogłyby funkcjonować. Chociaż sama inicjatywa ontologii SPAR wyszła ze środowiska badaczy reprezentujących nauki medyczne i biologiczne, to w założeniu miały one oferować nie tyle wyższy poziom ogólności, co pewnego rodzaju neutralność samej konceptualizacji. Zasadniczy trzon projektu SPAR tworzy osiem ontologii:

- (1) FRBR-aligned Bibliographic Ontology (FaBiO) – ontologia opisująca elementy uniwersum bibliograficznego oraz relacje zachodzące między nimi z punktu widzenia modelu FRBR;
- (2) Citation Typing Ontology (CiTO) – ontologia odwzorowująca typy cytowań bibliograficznych;
- (3) Bibliographic Reference Ontology (BiRO) – ontologia odwzorowująca strukturę bibliografii załącznikowej;
- (4) Citation Counting and Context Characterisation Ontology (C4O) – ontologia odwzorowująca odwołania bibliograficzne w treści publikacji naukowej w ujęciu bibliometrycznym;
- (5) Document Components Ontology (DoCO) – ontologia odwzorowująca strukturę publikacji naukowej;
- (6) Publishing Status Ontology (PSO) – ontologia opisująca status pracy naukowej na poszczególnych etapach procesu publikacji;
- (7) Publishing Roles Ontology (PRO) – ontologia opisująca role agentów (podmiotów) w procesie publikacji;
- (8) Publishing Workflow Ontology (PWO) – ontologia opisująca cały proces związany z publikowaniem pracy naukowej.

Wszystkie ontologie w projekcie SPAR zostały opracowane jako specyfikacje w języku OWL i są zarejestrowane w LOV. Ze względu na cel artykułu dalsze rozważania są poświęcone ontologii DOCO.

Głównym założeniem koncepcyjnym ontologii DOCO jest dekonstrukcja publikacji naukowej uwzględniająca jej elementy strukturalne i retoryczne (Constantin et al., 2016, 168). Reprezentacja publikacji naukowej jako dokumentu w ujęciu bibliograficznym jest tutaj zapewniona przez możliwość zastosowania ontologii FaBio, której poziom ekspresywności jest bardzo wysoki. Podstawą do opracowania konceptualizacji były wyniki uzyskane z:

- analizy piśmiennictwa, w tym wcześniejszych projektów z obszaru semantycznego publikowania;
- wytycznych dla autorów opracowanych przez wydawnictwa naukowe;
- nieformalnych wywiadów z naukowcami z różnych dyscyplin naukowych, które dotyczyły ich interpretacji elementów składowych publikacji naukowych oraz ich wykorzystania w pracy badawczej.

Autorzy ontologii przeprowadzili gruntowny przegląd stanu badań w zakresie analizy strukturalnej publikacji naukowych zarówno w ujęciu teorii gatunków, jak i analizy dyskursu. W dokumentacji DOCO i piśmiennictwie na jej temat występują liczne odwołania do prac naukowych dotyczących identyfikacji elementów strukturalnych i retorycznych publikacji naukowych. Pozwala to sądzić, że konceptualizacja domeny była poprzedzona zdobyciem gruntownej wiedzy dziedzinowej.

Ontologia DOCO ma również architekturę modułową. Obok własnego systemu wykładników klas i relacji odwzorowujących strukturę publikacji naukowej DOCO wykorzystuje

ontologii Discourse Elements Ontology (DEO)²⁰ oraz Pattern Ontology²¹. Pierwsza z nich jest rozszerzeniem ontologii SPAR i służy do modelowania elementów dyskursywnych publikacji naukowej, druga służy do odwzorowania schematów strukturalnych publikacji naukowych.

Strukturalna dekompozycja publikacji naukowej w ontologii DOCO zakłada trzy podstawowe elementy przejęte ze specyfikacji JATS:

- (1) Front matter – metadane na temat publikacji;
- (2) Body matter – wewnętrznie ustrukturyzowana zasadnicza treść publikacji;
- (3) Back matter – elementy aparatu pomocniczego publikacji.

Dla każdej z tych części opracowano szczegółową specyfikację elementów strukturalnych na wysokim poziomie szczegółowości (np. rozdział, sekcja, akapit, zdanie), uwzględniając dodatkowe obiekty tekstowe i graficzne występujące w jego treści (np. tabela, rysunek, formuła matematyczna) oraz ich wewnętrzną budowę z wykorzystaniem schematów strukturalnych Pattern Ontology.

Warstwa retoryczna publikacji naukowej została tutaj odwzorowana za pomocą ontologii DEO, która to z kolei została opracowana na podstawie istniejących, chociaż już nierozwijanych ontologii SALT Rhetorical Ontology (Groza, Handschuh, Möller, Decker, 2007) i The Ontology of Rhetorical Blocks (Ciccarese & Groza, 2011). Do podstawowych elementów retorycznych publikacji naukowej w ontologii DEO należą:

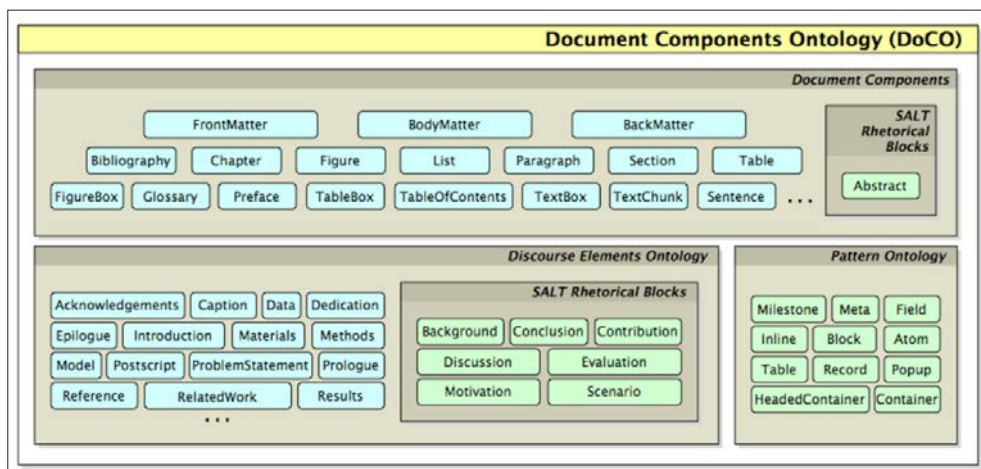
- (1) Odesłanie (ang. *Reference*) – odwołanie w tekście danej publikacji do jej elementu strukturalnego lub do innej publikacji;
- (2) Przypis bibliograficzny (ang. *Bibliographic Reference*) – odwołanie w postaci przypisu bibliograficznego;
- (3) Podpis (ang. *Caption*) – element tekstowy towarzyszący osadzonemu obiektowi (np. ilustracji);
- (4) Wstęp (ang. *Introduction*);
- (5) Materiał badawczy (ang. *Materials*) – wskazanie na dane badawcze lub inny typ zbioru informacji będącego przedmiotem dociekań. Element występuje najczęściej w części metodologicznej publikacji naukowej;
- (6) Metody (ang. *Methods*);
- (7) Wyniki badań (ang. *Results*);
- (8) Dyskusja (ang. *Discussion*);
- (9) Stan badań (ang. *Related Work*);
- (10) Dalsze badania (ang. *Future Work*).

Pełna specyfikacja tych elementów zawiera 47 klas reprezentujących pojęcia wskazujące na elementy retoryczne publikacji naukowej. Można zauważyć fakt, że DEO uwzględnia wszystkie elementy modelu IMRaD. Rysunek 3 przedstawia wizualizację najważniejszych klas ontologii DEO z podziałem na moduły.

Z przedstawionych analiz wyłania się obraz ontologii, którą cechuje duża ekspresywność, zarówno w odniesieniu do opisu warstwy strukturalnej, jak i retorycznej publikacji naukowej. Dotyczy to przede wszystkim dużej liczby klas zidentyfikowanych w tym modelu. Jeśli chodzi o sieć relacji występujących między tymi klasami, to ontologia DEO operuje jedynie relacją mereologiczną (z kilkoma podtypami) oraz niespecyfikowaną relacją asocjacyjną.

²⁰ <http://www.sparontologies.net/ontologies/deo>

²¹ <http://essepuntato.github.io/po/current/pattern.html>



Rys. 3. Architektura ontologii DEO.

Źródło: Constantin et al., 2016

Rozpatrując postawy epistemiczne, przyjęte w procesie konceptualizacji poprzedzającym konstrukcję ontologii, zauważyć można, że mamy tutaj do czynienia z deklaracją, iż przedmiotem poznania rzeczywistości była skodyfikowana wiedza naukowa na jej temat, co może świadczyć o podejściu realistyczno-pozytywistycznym i sytuować ten projekt po stronie krytycznego realizmu. Obecny w metodologii tego projektu czynnik badań nad użytkownikami pozwala sądzić, że poszukiwano również pewnego konsensu poznawczego w wymiarze społecznym, ale ze względu na brak informacji o przebiegu tego badania nie można wyprowadzać dalszych wiążących wniosków na płaszczyźnie epistemologicznej.

4. Wnioski

Przedstawione analizy prowadzą do trzech konstatacji, które odnoszą się do kwestii ontologicznych, interpretacyjnych oraz wdrożeniowych.

Po pierwsze, charakterystyka struktur pojęciowych leżących u podstaw trzech ontologii sieciowych, których celem jest reprezentacja artykułu naukowego na potrzeby semantycznego publikowania daje obraz modelu konceptualnego tego artefaktu naukowego, w którym przede wszystkim eksponuje się elementy pełniące określoną funkcję retoryczną. Raz są one nazywane elementami retorycznymi, innym razem dyskursywnymi. Tego rodzaju konceptualizacje nawiązują bezpośrednio do modelu IMRaD, ale w przeanalizowanych ontologiach znacznie go rozszerzają. W ontologii SciAnnotDoc oraz DOCO widoczna jest również próba odwzorowania struktury artykułu naukowego, która ma pełnić funkcje pomocnicze podczas identyfikacji elementów retorycznych, zapewniając kontekst ich obecności w treści samej publikacji. Nacisk na identyfikację elementów retorycznych skłania do refleksji, że są one postrzegane jako ważne elementy w procesie interpretacji treści publikacji naukowych i mogą być punktem odniesienia podczas automatycznej ekstrakcji informacji z tekstu. Niezależnie od przyjętych założeń metodologicznych, wydaje się, że najbardziej całościową postawą poznawczą, którą można przyjąć podczas konceptualizacji

tego uniwersum dyskursu jest identyfikacja trzech punktów widzenia:

- (1) Artykuł naukowy jako jednostka bibliograficzna;
- (2) Artykuł naukowy jako złożony strukturalnie artefakt naukowy;
- (3) Artykuł naukowy jako system elementów.

Taki punkt widzenia zapewnia z jednej strony odwzorowanie cech formalnych artykułu naukowego jako dokumentu, a z drugiej oferuje możliwość jego dekonstrukcji uwzględniającej elementy retoryczne w ramach określonych elementów strukturalnych. Taką perspektywę oferuje ontologia DOCO.

Po drugie, na płaszczyźnie interpretacyjnej podstawowym problemem w ocenie ontologii jest jej zrozumiałość (aspekt pragmatyczny) i przede wszystkim zbieżność modelu konceptualnego leżącego u jej podstaw z konceptualizacją fragmentu rzeczywistości, którą odwzorowuje i która jest współdzielona przez określoną społeczność. Uwagi o trudnościach z określeniem stopnia tego podobieństwa i tym samym wprowadzeniem kryteriów jakościowych do oceny ontologii pojawiają się w wielu publikacjach na temat ewaluacji ontologii (np. Brank et al., 2005; Gómez-Pérez, 2013; Suárez-Figueroa et al., 2012; Vrandečić, 2009). Zdaniem autora, tego typu relacja między ontologią a rzeczywistością, którą ona odwzorowuje jest zdeterminowana epistemologicznie. Ujawnienie przyjętych przez projektantów ontologii postaw epistemicznych dałoby szansę na pełniejszy obraz konsekwencji decyzji podjętych podczas procesu projektowania tych narzędzi. Próba ujawnienia postaw epistemicznych w przeanalizowanych trzech przypadkach polegała na analizie dokumentacji i piśmiennictwa na temat tych projektów i była oparta na ich rekonstrukcji na podstawie informacji o decyzjach metodologicznych. W przytoczonych przypadkach widać zarówno przyjmowanie postaw obiektywistycznych, jak i interpretatywnych, a także obecność determinantów o charakterze pragmatycznym. Z pewnością zarówno w przypadku modelowania obiegu informacji w nauce, jak i projektowania ontologii dziedzinowych potrzebna jest refleksja epistemologiczna w tym zakresie. Do ciekawych propozycji postaw epistemicznych, mogących mieć zastosowanie w tym przypadku, należą społeczno-pragmatyczny konstruktywizm (Bjeković et al., 2014) oraz pragmatyzm reprezentacyjny (Travers, 2011).

Przeprowadzona analiza piśmiennictwa i dokumentacji trzech ontologii pozwala również sformułować wniosek, że ich twórcy nie korzystają z bogatego dorobku naukowego innych dyscyplin, w których artykuł naukowy był przedmiotem badań w podobnym ujęciu. Dotyczy to przede wszystkim wykorzystania prac z teorii gatunków i analizy dyskursu. W pierwszym przypadku mowa o wkładzie Johna Swalesa (Swales, 2004) w badania nad artykułem naukowym z punktu widzenia teorii gatunków, w drugim, o licznych publikacjach Anity de Waard na temat identyfikacji elementów pragmatycznych i dyskursywnych w tego rodzaju tekstach (np. de Waard, 2007; Shum et al., 2010; de Waard et al., 2006) oraz jednego z najbardziej wyczerpujących ujęć problematyki struktury retorycznej publikacji naukowej w postaci rozprawy doktorskiej Frédérique Harmsze (2000). Tylko w przypadku prac na temat ontologii DOCO odnotowano odesłania do prac A. de Waard.

Po trzecie, ontologia jest narzędziem organizacji i reprezentacji wiedzy, które w założeniu ma funkcjonować w ramach określonego systemu informacyjnego. Wdrożenie tak ekspresywnych systemów organizacji wiedzy musi zakładać odpowiedź na pytanie, w jaki sposób będzie przebiegał proces indeksowania dokumentów z ich wykorzystaniem. Możliwe są tutaj przynajmniej trzy scenariusze. Pierwszy zakłada automatyczne indeksowanie zbioru dokumentów, drugi – wykorzystanie społeczności zgromadzonej wokół systemu

lub infrastruktury badawczej do tworzenia opisów zasobów kolekcji. W końcu, według trzeciego scenariusza, autor publikacji byłby włączony w ten proces i miałby zapewnione środki techniczne pozwalające mu na samodzielne sporządzenie opisu własnej publikacji podczas deponowania dokumentu do platformy publikacyjnej. Ostatni scenariusz jest obecnie rozwijany pod nazwą *autentycznego publikowania semantycznego* (ang. *genuine semantic publishing*) i opiera się na założeniu, że to właśnie autor jest jedynym wiarygodnym źródłem informacji na temat swojej publikacji (Kuhn & Dumontier, 2017, 146)

Pewną wątpliwość może budzić próba identyfikacji określonych elementów retorycznych w treści artykułu naukowego niezależnie od dyscypliny naukowej, którą ten artykuł reprezentuje. IMRaD, chociaż jest szeroko stosowanym narzędziem organizacji treści artykułu naukowego, *de facto* nie ma zastosowania we wszystkich dziedzinach wiedzy. Koncepcja semantycznego publikowania wyrosła na gruncie nauk medycznych i biologicznych, co spowodowało, że zaproponowane konceptualizacje, chociaż pretendowały do miana uniwersalnych, były w dużym stopniu zdeterminowane strukturą retoryczną tekstów właściwą dla tych nauk. Potrzebne są więc badania nad tego typu artefaktami naukowymi w naukach społecznych i humanistycznych, np. w ramach badań nad cyfrową humanistyką (np. Bartalesi & Meghini, 2016; Daquino & Tomasi, 2015) oraz kontroferata w postaci narzędzi prostszych do implementacji (np. Research Articles in Simplified HTML; Peroni et al., 2017)

Na koniec warto przywołać tezę Petera Miki (Mika, 2007), który podkreślił, że ontologie oraz Sieć Semantyczna, w ramach której funkcjonują są przeznaczone do przetwarzania przez aplikacje, ale proces ich tworzenia i zarządzania ma charakter społeczny. Tworzenie ontologii wymaga obecności czynnika społecznego, który polega na zdolności projektanta do przewidywania, w jaki sposób inni członkowie społeczności będą interpretowali postulaty znaczeniowe zapisane w danej ontologii.

Bibliografia

- ANSI/NISO (2018). *ANSI/NISO Z39.96–2019. JATS: Journal Article Tag Suite, version 1.2*. [online]. Baltimore, Md.: American National Standards Institute [01.06.2019]. https://groups.niso.org/apps/group_public/download.php/21030/ANSI-NISO-Z39.96-2019.pdf
- Bartalesi, V., Meghini, C. (2016). Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study. *Semantic Web* [online], 8(3), 385–394. <http://doi.org/10.3233/SW-150198>
- Bjeković, M., Proper, H. A., Sottet, J.-S. (2014). Embracing Pragmatics. In: E. Yu, G. Dobbie, M. Jarke, P. Sandeep (eds.), *Conceptual Modeling* [online]. *33rd International Conference, ER 2014 Atlanta, GA, USA, October 27–29, 2014 Proceedings* (431–444). Berlin: Springer. http://doi.org/10.1007/978-3-319-12206-9_37
- Brank, J., Grobelnik, M., Mladenčić, D. (2005). A Survey of Ontology Evaluation Techniques. In: M. Grobelnik, D. Mladenčić (eds.), *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)* [online], (166–169), [01.06.2019], <http://ailab.ijs.si/dunja/SiKDD2005/Papers/BrankEvaluationSiKDD2005.pdf>
- Ciccarese, P., & Groza, T. (2011). *Ontology of Rhetorical Blocks (ORB). W3C Interest Group Note 20 October 2011* [online], [01.06.2019], <https://www.w3.org/TR/hcls-orb/>
- Constantin, A., Peroni, S., Pettifer, S., Shotton, D., Vitali, F. (2016). The Document Components Ontology (DoCO). *Semantic Web* [online], 7(2), 167–181. <http://doi.org/10.3233/SW-150177>
- Daquino, M., & Tomasi, F. (2015). Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects. In: Garoufallou E., R. Hartley, & P. Gaitanou

- (eds.), *Metadata and Semantics Research [online]*. MTSR 2015. *Communications in Computer and Information Science*, vol 544 (424–436). Berlin: Springer. http://doi.org/10.1007/978-3-319-24129-6_37
- Fernández-López, M., Gómez-Pérez, A. (2002). Overview and Analysis of Methodologies for Building Ontologies. *The Knowledge Engineering Review [online]*, 17(2), 129–156. <http://doi.org/10.1017/S0269888902000462>
- Gerstein, M., Seringhaus, M., Fields, S. (2007). Structured Digital Abstract Makes Text Mining Easy. *Nature [online]*, 447(7141), 142–142. <http://doi.org/10.1038/447142a>
- Gómez-Pérez, A. (2013). Ontology Evaluation. In: S. Staab & R. Studer (eds.), *Handbook on Ontologies* (251–274) [online]. Berlin, Heidelberg: Springer. http://doi.org/10.1007/978-3-540-24750-0_13
- Groza, T., Handschuh, S., Möller, K., Decker, S. (2007). SALT – Semantically Annotated LaTeX for Scientific Publications. In: E. Franconi, M. Kifer, W. May (eds.), *The Semantic Web: Research and Applications* (518–532) [online]. Berlin, Heidelberg: Springer. http://doi.org/10.1007/978-3-540-72667-8_37
- Guarino, N., Oberle, D., Staab, S. (2009). What Is an Ontology? In: S. Staab & R. Studer (eds.), *Handbook on Ontologies [online]*. Berlin, Heidelberg: Springer Berlin Heidelberg. <http://doi.org/10.1007/978-3-540-92673-3>
- Guizzardi, G. (2005, March 1). *Ontological Foundations for Structural Conceptual Models [online]*. CTIT, Centre for Telematics and Information Technology [01.06.2019], http://doc.utwente.nl/50826/1/thesis_Guizzardi.pdf
- Harmsze, F. A. P. (2000). *A Modular Structure for Scientific Articles in an Electronic Environment [online]*. University of Amsterdam [01.06.2019], <https://dare.uva.nl/search?arno.record.id=78293>
- Hjørland, B. (2003). Fundamentals of Knowledge Organization. Knowledge Organization. *Knowledge Organization*, 30(2), 87–111.
- Klein, H. K., Hirschheim, R. A. (1987). A Comparative Framework of Data Modelling Paradigms and Approaches. *The Computer Journal [online]*, 30(1), 8–15. <http://doi.org/10.1093/comjnl/30.1.8>
- Kuhn, T., Dumontier, M. (2017). Genuine Semantic Publishing. *Data Science [online]*, 1, 1–16. <http://doi.org/10.3233/DS-170010>
- Liakata, M., Thompson, P., Waard, A. de, Nawaz, R., Maat, H. P., Ananiadou, S. (2012). A Three-Way Perspective on Scientific Discourse Annotation for Knowledge Extraction. In: A. Van Den Bosch, H. Shatkey (eds.), *ACL '12 Proceedings of the Workshop on Detecting Structure in Scholarly Discourse [online]*, (37–46), [01.06.2019]. Jeju Island, Korea: Association for Computational Linguistics. <https://www.aclweb.org/anthology/papers/W/W12/W12-4305/>
- Lizzi, V. (2017). Implementation of JATS at Taylor & Francis. In: *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2017 [online]*. Bethesda: National Center for Biotechnology Information [01.06.2019], <https://www.ncbi.nlm.nih.gov/books/NBK425705/>
- Mazzocchi, F. (2017). Knowledge Organization System (KOS). In: B. Hjørland (ed.), *Encyclopedia of Knowledge Organization [online]*. ISKO [01.06.2019], <http://www.isko.org/cyclo/kos>
- McGuinness, D. L., Noy, N. F. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology [online]*. Knowledge Systems Laboratory Stanford University [01.06.2019], http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- Meister, V. G. (2017). Towards a Knowledge Graph for a Research Group with Focus on Qualitative Analysis of Scholarly Papers. In: D. Garijo, W. R. van Hage, T. Kauppinen, T. Kuhn, J. Zhao (eds.), *Enabling Open Semantic Science. Proceedings of the First Workshop on Enabling Open Semantic Science co-located with 16th International Semantic Web Conference (ISWC 2017) [online]*, 71–76, [01.06.2019], <http://ceur-ws.org/Vol-1931/paper-10.pdf>
- Mika, P. (2007). Ontologies Are Us: A Unified Model of Social Networks and Semantics. *Journal of Web Semantics*, 5(1), 5–15.
- Mitterer, J. (2017). Conceptual Modeling: Philosophical Considerations. In: H. C. Mayr & G. Guizzardi (eds.), *Conceptual Modeling 36th International Conference, ER 2017 Valencia, Spain, November 6–9, 2017 Proceedings* (15). Berlin, Heidelberg: Springer.

- Niehaves, B., Becker, J. (2006). Epistemological Perspectives on Design Science in IS Research. In: *AMCIS 2006 Proceedings* [online], (258–261), [01.06.2019]. San Diego: Idea Group. <https://aisel.aisnet.org/amcis2006/430>
- Penev, L. (2017). From Open Access to Open Science from the Viewpoint of a Scholarly Publisher. *Research Ideas and Outcomes* [online], 3, e12265. <http://doi.org/10.3897/rio.3.e12265>
- Pérez, A. G., Carmen, M., Figueroa, S. De, Villazón, B. (2008). *NeOn Methodology for Building Ontology Networks : Ontology Specification* [online]. NeOn Project [01.06.2019], <http://neon-project.org/nw/Deliverables.html>
- Peroni, S., Osborne, F., Di Iorio, A., Nuzzolese, A. G., Poggi, F., Vitali, F., Motta, E. (2017). Research Articles in Simplified HTML: A Web-first Format for HTML-based Scholarly Articles. *PeerJ Computer Science* [online], 3, e132, <http://doi.org/10.7717/peerj-cs.132>
- Recker, J., Niehaves, B. (2008). Epistemological Perspectives on Ontology-based Theories for Conceptual Modeling. *Applied Ontology* [online], 3(1–2), 111–130, <http://doi.org/10.3233/AO-2008-0045>
- Ribbert, M., Niehaves, B., Dreiling, A., Holten, R. (2004). An Epistemological Foundation of Conceptual Modeling [online]. In: *ECIS 2004 Proceedings. Paper 113*, (August), (4232–4242). [01.06.2019], <https://aisel.aisnet.org/ecis2004/113/>
- Ribaupierre, H., de (2014). *Precise Information Retrieval in Semantic Scientific Digital Libraries* [online]. Archive ouverte UNIGE [01.06.2019], <https://archive-ouverte.unige.ch/unige:43165>
- Ribaupierre, H., de, Falquet, G. (2018). Extracting Discourse Elements and Annotating Scientific Documents Using the SciAnnotDoc Model: A Use Case in Gender Documents. *International Journal on Digital Libraries* [online], 19(2–3), 271–286, <http://doi.org/10.1007/s00799-017-0227-5>
- Shotton, D. (2009). Semantic Publishing: The Coming Revolution in Scientific Journal Publishing. *Learned Publishing* [online], 22(2), 85–94, <http://doi.org/10.1087/2009202>
- Shum, S., Clark, T., Waard, A., de (2010). Scientific Discourse on the Semantic Web: A Survey of Models and Enabling Technologies [online], [01.06.2019], <http://www.semantic-web-journal.net/content/scientific-discourse-semantic-web-survey-models-and-enabling-technologies>
- Studer, R., Benjamins, V. R., Fensel, D. (1998). Knowledge Engineering: Principles and Methods. *Data & Knowledge Engineering* [online], 25(1–2), 161–197, [http://doi.org/10.1016/S0169-023X\(97\)00056-6](http://doi.org/10.1016/S0169-023X(97)00056-6)
- Suárez-Figueroa, M. C., Gómez-Pérez, A., Motta, E., Gangemi, A. (eds.). (2012). *Ontology Engineering in a Networked World* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg. <http://doi.org/10.1007/978-3-642-24794-1>
- Swales, J. M. (2004). *Research Genres: Explorations and Applications* [online]. Cambridge: Cambridge University Press, <http://doi.org/10.1017/CBO9781139524827>
- Travers, M. (2011). Politics and Pragmatism in Scientific Ontology Construction. *Inconsistency Robustness* [online], (May), 1–33 [01.06.2019], <http://www.ai.sri.com/~travers/onto-revised.pdf>
- Vicente-Saez, R., Martinez-Fuentes, C. (2018). Open Science Now: A Systematic Literature Review for an Integrated Definition. *Journal of Business Research* [online], 88, 428–436. <http://doi.org/10.1016/j.jbusres.2017.12.043>
- Vrandečić, D. (2009). Ontology Evaluation. In: Staab S., Studer R. (eds.) *Handbook on Ontologies* [online]. Berlin, Heidelberg: Springer. http://doi.org/10.1007/978-3-540-92673-3_13
- Waard, A., de (2007). A Pragmatic Structure for Research Articles. In: *Proceedings of the 2nd international conference on Pragmatic web – ICPW '07* (83–89) [online]. New York, USA: ACM Press. <http://doi.org/10.1145/1324237.1324247>
- Waard, A., de, Breure, L., Kircz, J. G., Oostendorp, H. Van. (2006). Modeling Rhetoric in Scientific Publications. In: V. P. Guerrero Bote (ed.), *Current Research in Information Sciences and Technologies. International Conference on Multidisciplinary Information Sciences and Technologies InSciT2006 Badajoz, Spain: 25–28 October 2006* (352–356). Bajados, Spain: Open Institute of Knowledge.
- Zins, C. (2004). Knowledge Organization: An Epistemological Perspective. *Knowledge Organization*, 31(1), 49–54.

Deconstructing the Scholarly Paper. Ontologies for Semantic Publishing

Abstract

Purpose/Thesis: The aim of this paper is to study three ontologies developed in the domain of Semantic Publishing for describing academic papers – SciAnnotDoc, Scholarly Papers Vocabulary with Focus on Qualitative Analysis, Document Components Ontology.

Approach/Methods: The study follows the method of ontology assessment and is based on the interpretation of meaning postulates. The study of ontologies is based on the schema: ontology's scope and domain, considerations of the ontology's context, ontological premises; furthermore, it attempts to identify of epistemic stance taken during the process of construction.

Results and conclusions: The results of the study show that conceptual structures behind these the three ontologies first of all expose rhetorical or discursive elements of scholarly paper. In all of three cases, IMRaD was not the first choice for structuring the content of a publication. It was not possible to fully reveal epistemic stances taken with the regards to the three ontologies. However, when stance was identified, it was possible to discern both objectivist and interpretative approaches as well as pragmatic determinants.

Originality/Value: Conceptual modeling, which is one of the initial stages of the ontology design process is affected by the epistemological approach, i.e. the attitude of the ontologists towards the reality, as they try to represent its part by the means of ontology. Revealing these epistemic stances is crucial for understanding the context of meaning postulates in these knowledge organization systems.

Keywords

Conceptual modelling. Epistemic stances. Ontologies. Scholarly papers. Semantic publishing.

Dr MARCIN ROSZKOWSKI jest adiunktem w Katedrze Informatologii na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego, członkiem International Society for Knowledge Organization oraz Komitetu ds. Ontologii w projekcie DBpedia. Jego zainteresowania naukowe obejmują problematykę organizacji wiedzy i reprezentacji informacji w środowisku sieciowym ze szczególnym uwzględnieniem modelowania konceptualnego systemów informacyjnych oraz metadanych i ontologii sieciowych. Najważniejsze publikacje: Organizacja informacji i wiedzy (W: W. Babik, red., Nauka o informacji, 2016; z B. Sośnińską-Kalatą); The Role of Digital Libraries as Virtual Research Environments for the Digital Humanities (In: J. A. C. Guimarães, S. Oliveira Milani, & V. Dodebei, eds., Advances in Knowledge Organization, 2016, z W. Mustafa El Hadi); Kartoteka haseł wzorcowych jako usługa sieciowa – automatyczna identyfikacja nazw osobowych z wykorzystaniem kartoteki VIAF (W: J. Woźniak-Kasperek & J. Franke, red. Bibliografia – teoria, praktyka, dydaktyka, 2016).

Kontakt z autorem:

m.rozkowski@uw.edu.pl

Katedra Informatologii

Wydział Dziennikarstwa, Informacji i Bibliologii

Uniwersytet Warszawski

ul. Nowy Świat 69

00-046 Warszawa