

w niczym bardziej pierwotnym), to dzięki tej sytuacji teoretycznej otwiera się droga ku prawdziwej transcendencji – transcendencji tego, co jest absolutnie anonimowe i nigdy niepoznawalne (na razie nawet w swoim istnieniu).

Wit Wawrzyniak
doktorant w Katedrze Epistemologii
na Wydziale Filozofii KUL

Krzysztof Goczyła, *Ontologie w systemach informatycznych*, Warszawa: Akademicka Oficyna Wydawnicza EXIT 2011, ss. 310. ISBN: 978-83-60434-88-8.

Szybki rozwój technologii komputerowych, z Internetem na czele, oraz ciągłe zmiany w społeczeństwie, gdzie od dłuższego już czasu informacja, a od niedawna również wiedza stają się coraz ważniejszym towarem, sprawiają, że dotychczasowe „mechaniczne” metody przetwarzania danych przestają być zadowalające. Przetwarzaniem danych i informacji zajmuje się gałąź informatyki nazywana inżynierią informacji. Bada ona metody przechowywania i przetwarzania informacji pod względem ich użyteczności dla użytkowników i szuka takich, które najlepiej spełniają oczekiwania użytkowników. W ostatnim czasie pojawiła się nowa poddziedzina inżynierii informacji nazwana inżynierią ontologiczną, która wynosi przetwarzanie informacji na wyższy i bliższy użytkownikom poziom – poziom znaczenia. Na skutek badań prowadzonych m.in. w inżynierii ontologicznej powstał pomysł Semantycznego Internetu (*Semantic Web*), którego celem jest opisanie znajdujących się w Internecie danych za pomocą znaczników nadających im sens (znaczenie). Tematem recenzowanej książki są technologie stosowane w ramach Semantycznego Internetu, które służą realizacji idei semantycznego opisywania danych w praktyce. Jakkolwiek inżynieria ontologiczna jest działem informatyki, to czerpie ona liczne inspiracje z systemów filozoficznych, zwłaszcza w przypadku tzw. ontologii fundacjonistycznych, które starają się opisać najbardziej podstawowe kategorie do reprezentacji świata i na których podstawie tworzone są ontologie bardziej specjalistyczne (dziedzinowe).

Recenzowana książka składa się z dziewięciu rozdziałów, które we „Wstępie” sam Autor nieformalnie (podział ten nie ma odzwierciedlenia w strukturze książki) podzielił na dwie części. Część pierwsza (rozdziały od pierwszego do piątego) jest bardziej praktyczna, omawia narzędzia stosowane w inżynierii ontologicznej do reprezentowania wiedzy. W części drugiej (rozdziały od szóstego do dziewiątego) zostały przedstawione teoretyczne podstawy koncepcji omawianych w pierwszej części.

Właściwą część książki Autor poprzedza wstępem, w którym prezentuje doniosłość zagadnień poruszanych w książce, omawia strukturę pracy i zamieszcza zwyczajowe

podziękowania. W rozdziale pierwszym („Dane, informacje i wiedza”) Goczyła przedstawia kilka podstawowych pojęć występujących w inżynierii informacji. Do pojęć tych zalicza się m.in. pojęcie danych, które przybliża za pomocą pojęcia typu danych, rozumianego jako para $T = (W, O)$, „gdzie W jest zbiorem wartości dopuszczalnych dla tego typu (zwanym też niekiedy dziedziną typu), a O jest zbiorem operacji, które można wykonywać na tych wartościach” (s. 15). Tego typu pojęcie danych jest blisko związane z pojęciem danych występującym w systemach komputerowych. W tym rozumieniu daną może być liczba 5, napis „Andrzej”, data 11-09-2001 itp. Na podstawie pojęcia danych buduje się pojęcie informacji. Czym, według Goczyły, jest informacja? Informacja to dane powiązane za pomocą jakichś relacji. Typowym przykładem, jaki przywołuje Goczyła w celu zaprezentowania użycia informacji w tym znaczeniu, są relacyjne bazy danych, gdzie dane „mają swoje określone znaczenie, nadane im przez związki” (s. 18). Dalej Autor przechodzi do omówienia pojęcia wiedzy. Najpierw rozpatruje potoczne rozumienia tego pojęcia, by następnie omówić pojęcie wiedzy komputerowej: „[w]iedza komputerowa jest to zbiór informacji zapisanych w pamięci komputera wraz ze zdolnością komputera do samodzielnego poszerzenia tego zbioru drogą wnioskowania” (s. 20). Jak widać, elementem, który rozróżnia informacje od wiedzy, jest możliwość wyprowadzania (wnioskowania) nowej wiedzy z już zdobytego zbioru informacji.

W drugim rozdziale („Wprowadzenie do ontologii”) Goczyła opisuje, czym są ontologie inżynieryjne. Wychodzi od jednej z definicji ontologii filozoficznej, ale nie zatrzymuje się na niej ani nie stara się porównywać ontologii filozoficznych z ontologiami inżynieryjnymi, lecz przechodzi bezpośrednio do pojęcia ontologii, jakie stosuje się powszechnie w ramach Semantycznego Internetu, czyli do definicji Grubera:

Ontologia jest formalną, jawną specyfikacją wspólnej konceptualizacji.

Goczyła omawia kolejno znaczenia poszczególnych pojęć wchodzących w skład tej definicji. Dowiadujemy się więc, że „formalność” oznacza bycie zapisanym w formacie zrozumiałym dla komputera. Choć warunek ten nie zawsze jest spełniony, to został on umieszczony w definicji, by podkreślić dążenia twórców Semantycznego Internetu, którzy chcą, żeby komputery w Internecie mogły wymieniać się wiedzą bez interwencji użytkownika. W tym celu potrzebne są standardy wymiany danych, które zostały omówione w dalszych rozdziałach. Wracając do definicji Grubera, słowo „jawną” oznacza, że ontologia ma prezentować wiedzę całościową, nie zakłada się istnienia bytów niejawnych. „Wspólna” znaczy tyle, że ontologia powinna ujmować wiedzę powszechnie dostępną, wspólną dla grupy zainteresowanych podmiotów (np. dla osób, które mają w przyszłości korzystać z pojęciowych rozstrzygnięć poczynionych w danej ontologii). Słowo „konceptualizacja” znaczy, że ontologia jest modelem pewnej realnej dziedziny ze świata (oczywiście można próbować konceptualizować cały świat, ale zadanie takie wydaje się niemożliwe do realizacji z przyczyn praktycznych).

Dalej Autor prezentuje klasyfikację ontologii. Kryterium podziału jest poziom formalizacji ontologii. Najślabszą pod względem formalizacji ontologią jest predefiniowane słownictwo, następnie słowniki, tezaury oraz taksonomie. Wszystkie te odmiany ontologii są przez Autora określone jako nieformalne. Do ontologii formalnych Goczyła zalicza ontologie oparte na danych oraz ontologie oparte na logice. Dalej w książce są omawiane przede wszystkim ontologie formalne, a zwłaszcza ontologie oparte na logice. Następnie Autor wspomina jeszcze o innych kryteriach podziału ontologii, ale niestety jedynie pobieżnie¹. Na koniec tego rozdziału Goczyła prezentuje (w postaci grafów) dla przykładu dwie proste ontologie oraz wylicza szereg odnośników internetowych do już istniejących ontologii.

W rozdziale trzecim („Ramy i sieci semantyczne”) zaprezentowano koncepcję ram Minsky’ego. Koncepcja ta została opisana przez Autora jako „idea, która wywarła największy wpływ na rozwój ontologicznych metod reprezentacji wiedzy, a także na rozwój obiektowych metod w inżynierii oprogramowania” (s. 35). Niestety, Autor nie popiera tego stwierdzenia żadnym dowodem, a jedynie stwierdza, że „uważa się [tak – R.T.]” (s. 35). Nawet jeżeli jest prawdą, że ramy Minsky’ego miały tak duży wpływ na współczesne metody stosowane w inżynierii wiedzy, to i tak budzi wątpliwość fakt prezentowania ich w książce, która ma być wprowadzeniem do tematu. Wydaje mi się bowiem, że dzisiaj ramy Minsky’ego nie są już popularnym narzędziem i zostały zastąpione przez nowsze metody. Z tego też powodu nie będę prezentował ustaleń Goczyły na temat ram Minsky’ego, tym bardziej że koncepcja ta nie jest używana dalej w książce. Kolejną metodą reprezentacji wiedzy, jaką przedstawia Goczyła, są sieci semantyczne. U podstaw tej koncepcji leży próba stworzenia modelu reprezentacji wiedzy możliwie bliskiego sposobowi funkcjonowania ludzkiego mózgu. Jest to model asocjacyjny, w którym pojęcia łączone są między sobą za pomocą relacji. Połączenia te pozwalają lepiej zrozumieć pojęcie, od którego wychodzą. Na grafie sieci semantyczne reprezentuje się za pomocą dwóch różnych typów elementów: węzłów i łuków. Przy tym węzłami mogą być zarówno pojęcia abstrakcyjne (klasy), jak i konkretne obiekty. Łuki natomiast wyrażają relację łączącą przedmioty.

Kolejny rozdział, czwarty („Resource Description Framework (RDF)”), zawiera opis języka RDF, który służy do zapisu metadanych (czyli danych służących do opisu innych danych). RDF jest podstawową cegiełką, z której złożony jest Semantyczny Internet, oraz punktem wyjścia dla szeregu innych technologii, które dają większe możliwości opisu danych. Jednym z przykładów języka, który został nabudowany na RDF, jest, omawiany w następnym rozdziale, język OWL. Rozdział dotyczący RDF-a zawiera podstawowe informacje o RDF-ie, prezentuje składnię tego języka oraz opis języka SPARQL, który służy do tworzenia zapytań (na podobnej zasadzie, co SQL dla relacyjnych baz danych) do baz wiedzy z wykorzystaniem idei pochodzących z RDF-a. W recenzji skupię się

¹ Więcej informacji o różnych podziałach ontologii Czytelnik może znaleźć w pracy: P. G a r - b a c z, R. T r y p u z, *Ontologie poza ontologią*, Lublin: Wydawnictwo KUL 2012.

jedynie na podaniu kilku podstawowych informacji, do czego służy RDF i jak wyglądają najprostsze przypadki użycia. Podstawowym zadaniem, jakie postawiono przed językiem opisu metadanych RDF jest opis danych istniejących w Internecie, wzbogacenie ich o znaczenie. Informacje o znaczeniu pozwalają komputerowi i człowiekowi ocenić na przykład, czy dana informacja wiąże się z poszukiwanym tematem.

Goczyła omawia dwie podstawowe zasady, które leżą u podstaw standardu RDF. Pierwsza z nich mówi, że wszystko jest zasobem, co znaczy, że wszystko może być identyfikowane przez adres internetowy, który pozwala jednoznacznie identyfikować dany zasób. Jeżeli nadamy jakiejś osobie (która jest zasobem) adres internetowy, to za pomocą tego adresu możemy przypisać tej osobie imię, nazwisko, adres, afiliację, datę urodzenia itd. Jak widać, zasób jest tu rozumiany bardzo szeroko i może się odnosić zarówno do przedmiotów istniejących fizycznie (domy, samochody, ludzie, zwierzęta) oraz nieistniejących fizycznie (bohater książki, postać z mitologii). Możemy opisywać klasy rzeczy oraz poszczególne indywidua. Druga zasada mówi, że każdy zasób należy opisywać za pomocą tzw. trójek RDF. Każda trójka to proste zdanie, które składa się z podmiotu, orzeczenia i dopełnienia. Każdym z tych elementów odnosi się do jakiegoś zasobu (za dopełnienie możemy podstawić również tzw. literał, czyli daną reprezentowaną przez ciąg znaków). Załóżmy, że mamy zdefiniowaną relację-zasób „urodzony” oraz zasób reprezentujący Jana Kowalskiego. Za ich pomocą możemy określić, że Jan Kowalski urodził się 13.06.1950:

janKowalski urodzony „13.06.1950”.

Powyższy uproszczony zapis nie jest poprawnym zapisem w RDF-ie, ale ma służyć jedynie prezentacji idei. Pomimo prostoty trójek RDF-owych podejście to daje duże możliwości opisu danych.

Rozdział piąty („Web Ontology Language (OWL)”) opisuje wspomniany już język OWL, który rozszerza poprzedzające go języki opisu danych i jest obecnie najpopularniejszym językiem do zapisu ontologii stosowanej. Podobnie jak RDF, OWL jest standardem promowanym przez organizację W3C (World Wide Web Consortium), która opiekuje się szeregiem standardów internetowych, a na której czele stoi Tim Berners-Lee, twórca (pierwszej wersji) WWW. Jako podstawę dla OWL przyjęto logikę opisową (deskrypcyjną), która jest rozstrzygalnym podzbiorem logiki predykatów pierwszego rzędu. Logika opisowa została przez Goczyłę opisaną w części bardziej teoretycznej (rozdziały szósty i siódmy). Pierwsza wersja standardu języka OWL została stworzona w 2004 r. i oznaczona numerem 1.0. W książce Autor omawia standard w wersji 2.0, który ukazał się w 2009 r. i znacząco usprawnia język OWL. Składnia języka OWL, jak już wspominałem, opiera się na składni języka RDF. Po informacjach podstawowych, Goczyła przechodzi do szczegółowego opisu języka OWL. Całe zagadnienie dzieli na dwie części: „konstrukcje podstawowe” i „konstrukcje zaawansowane”. W „konstrukcjach podstawowych” omawia takie zagadnienia, jak stwierdzenie, że coś jest klasą (Class-

Assertion), relacja bycia podklasą, relacja równoważności klas, rozłączność klas itd. W „konstrukcjach zaawansowanych” zaś Autor opisuje aksjomaty dotyczące właściwości, typy danych i atrybuty itp. Nie będę tutaj referował tych informacji ze względu na szerokość tematu. Trzeba jednak stwierdzić, że narracja Goczyły jest jasna, precyzyjna i zawiera szereg przykładów, które ułatwiają zrozumienie nawet trudniejszych zagadnień.

Rozdział szósty („Logika opisowa (DL)”) rozpoczyna część bardziej teoretyczną, gdzie wyjaśniane są koncepcje służące jako podstawa teoretyczna dla wcześniej omawianych języków i standardów. Fakt, że język OWL został oparty na pewnym podzbiorze logiki pierwszego rzędu, który jest rozstrzygalny, pozwala domyślać się, że jedną z cech ontologii zapisanych w języku OWL jest możliwość wyprowadzania z nich nowych wniosków na podstawie wprowadzonych wcześniej danych. Ontologie OWL są zapisywane w różnych notacjach, z których każda może być odczytywana przez komputer i tak zwane silniki wnioskujące, które pozwalają na zautomatyzowanie wnioskowań. Logika opisowa ma wiele dialektów, z których każdy ma nieco inną moc inferencyjną i stopień ekspresji. W tym rozdziale Goczyła opisuje różnice między logiką opisową a rachunkiem predykatów pierwszego rzędu, podstawowe elementy składniowe logiki opisowej, a następnie wprowadza poszczególne dialekty logiki deskrypcyjnej. W kolejnym podrozdziale Autor zwraca uwagę na podstawowe problemy wnioskowania, by dalej przejść do opisu możliwości włączenia do logiki opisowej metod wnioskowania opartych na regułach.

W rozdziale siódmym („Algorytmy wnioskowania z ontologii DL”) Autor omawia trzy metody wnioskowania stosujące się do ontologii. Żeby wykorzystać jedną z najważniejszych zalet ontologii, czyli możliwość wnioskowania nowej wiedzy przez komputer, musimy stworzyć algorytm (metodę) wnioskowania, którą będzie można zaimplementować jako program komputerowy. Goczyła omawia trzy takie metody: subsumcję strukturalną, algorytm tablicowy oraz kartograficzną metodę analizy ontologii. Każda z tych metod ma zalety i wady. Najprostszą z nich jest metoda subsumcji strukturalnej, w której za prostotę koncepcyjną i implementacyjną przychodzi płacić ograniczoną stosownością. Algorytm ten opiera się na badaniu struktury konceptów. Druga metoda, algorytm tablicowy, jest najszerzej stosowana, ale trudniejsza do zrozumienia i implementacji. Opiera się na idei badania spełnialności, czyli istnienia chociaż jednego modelu dla danego zbioru aksjomatów i asercji. Trzecia metoda, metoda tzw. kartograficznej analizy ontologii, opiera się na tworzeniu „mapy konceptów” dla ontologii. Goczyła dość szczegółowo omawia każdą z tych metod i wskazuje ich zalety oraz wady.

W rozdziale ósmym („Bazy wiedzy a bazy danych”) Autor omawia różnice występujące między bazami wiedzy a bazami danych. Bazy danych, których najczęściej spotykana realizacją są relacyjne bazy danych, różnią się od baz wiedzy przede wszystkim tym, że podlegają założeniu o zamkniętym świecie. W bazach danych przyjmuje się założenie o zamkniętości świata, natomiast w bazach wiedzy przyjmuje się, że świat jest otwarty. Co to znaczy, że świat jest otwarty? „Zgodnie z tym założeniem, informacje zapisane w bazie wiedzy traktowane są jak niekompletne. Podczas wnioskowania baza wiedzy bierze pod uwagę nie tylko te informacje, które są w niej jawnie zapisane, ale także

uwzględnia możliwość prawdziwości innych, niezapisanych w bazie wiedzy informacji, które mogą wpłynąć na wynik wnioskowania” (s. 242). Założenie o otwartości świata w bazach wiedzy łączy się ściśle z ideą Semantycznego Internetu, ponieważ żaden użytkownik w Internecie nie może opisać wszystkich indywiduów danej klasy (np. klasy ludzi). Bazy danych przyjmują, odwrotnie, założenie o zamkniętym świecie: istnieje tylko to, co jest zapisane w bazie danych. Goczyła podaje dalej kilka przykładów, które pozwalają zrozumieć różnicę między tymi dwoma podejściami. W kolejnym podrozdziale omawia zagadnienie domykania świata oraz operator epistemologiczny K, który służy do domykania. Czasem może się bowiem okazać, że założenie o otwartym świecie jest dla naszych celów za mocne. W ostatnim podrozdziale Goczyła referuje zagadnienie monotoniczności w kontekście założenia o otwartości świata.

Ostatni rozdział, dziewiąty („Elementy inżynierii wiedzy”), zawiera wybrane praktyczne zagadnienia związane z tworzeniem ontologii. Goczyła przedstawia w nim typowe błędy, które pojawiają się podczas tworzenia ontologii, oraz ogólne wskazówki dotyczące budowania ontologii. Omawiane błędy są ilustrowane na przykładach, co ułatwia zrozumienie tematu.

Na końcu rozdziałów od czwartego do dziewiątego zamieszczone są zadania, które pozwalają czytelnikowi lepiej zrozumieć omawiane zagadnienia. Do większości zadań Autor dał rozwiązania na końcu książki. Książka zawiera kilka dodatków, które uzupełniają zasadniczą treść pracy. W dodatku A zostały opisane konstrukcje OWL 2 i odpowiadające im konstrukcje logiki opisowej. W dodatku B opisano niektóre użyteczne w projektowaniu ontologii zależności zapisane w języku logiki opisowej. Dodatek C zawiera zestawienie związane z różnymi dialektami logiki opisowej oraz dopuszczalnymi w nich wyrażeniami logiki opisowej. Dodatek D zawiera słowniczek angielsko-polski, który ułatwi czytelnikowi lekturę literatury przedmiotu w języku angielskim. Książkę zamyka bibliografia, w której przy niektórych pozycjach znajdują się krótkie komentarze Goczyły, oraz skorowidz.

Książka Goczyły to pierwsza praca poświęcona inżynierii ontologicznej napisana w języku polskim. I choć można powiedzieć, że dopiero przeciera dziewicze szlaki, to jej poziom robi duże wrażenie. Logiczne uporządkowanie tematu, konkretne, rzeczowo napisane rozdziały, brak zbędnego teoretyzowania, poprawny język (co nie jest znowu tak częste w książkach technicznych) – to podstawowe zalety recenzowanej pracy, które pozwalają mi polecić niniejszą książkę wszystkim zainteresowanym tematem ontologii stosowanej.

Rafał Trójczak
doktorant w Katedrze Podstaw Informatyki
na Wydziale Filozofii KUL