

Uniwersytet Szczeciński \* Wydział Humanistyczny  
Instytut Filozofii

Kazimierz Pogoda

hshsce@gmail.com

Zastosowania semantyki w kontekście  
współczesnych sieci informatycznych

Praca magisterska  
napisana pod kierunkiem  
prof. dra hab. Maxa Urchsa  
kod dziedziny: 12  
(*cognitive science*)

Szczecin 2004

---

## Od autora

Niniejszą pracę magisterską obroniłem w czerwcu 2004 roku. Pomimo iż od tamtego czasu minęło kilka lat, temat z każdym dniem zdaje się zyskiwać na aktualności. Postanowiłem zatem udostępnić tekst w Internecie, licząc na to, że wysiłek który włożyłem w opracowanie go posłuży czemuś więcej, niż rozwijanie hodowli roztoczy kurzu domowego, tudzież zwiększanie różnorodności domen magnetycznych dysku twardego mego komputera.

Tekst ma charakter filozoficzny. Praktyk szukający tu konkretnych informacji może się srogo zawieść. Z drugiej strony w pracy zarysowane zostaje coś na kształt „programu badawczego”. Gdyby ktoś chciał kontynuować te badania, z góry oferuję swą pomoc. W tej, jak i wielu innych sprawach, można kontaktować się ze mną pod adresem email:

[hshsce@gmail.com](mailto:hshsce@gmail.com)

W wielu miejscach tekstu wprowadziłem korekty, a dalszych poprawek można też spodziewać się w przyszłości. Nie mam jednak zamiaru rozwijać tematu dalej w tej formie. Nie znaczy to, iż straciłem zainteresowanie samym zagadnieniem obecności semantyki w Internecie. Najświeższe informacje odnośnie moich projektów dostępne są na stronie:

<http://blog.xemantic.com/>

Szczecin, 4 stycznia 2008

### **Oświadczenie**

Oświadczam, że załączoną pracę magisterską, kończącą studia napisałem samodzielnie. Oznacza to, że przy pisaniu pracy poza niezbędnymi konsultacjami, nie korzystałem z pomocy innych osób, a w szczególności nie zlecałem opracowania rozprawy lub jej części innym osobom, ani nie odpisywałem tej rozprawy lub jej części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Kazimierz Pogoda

## Spis treści

<b>1</b>	<b>Wstęp</b>	<b>8</b>
<b>2</b>	<b>Internet</b>	<b>9</b>
2.1	Zakres znaczeniowy pojęcia . . . . .	10
2.2	Geneza . . . . .	12
2.3	<i>Warstwa transportowa</i> , a przekazywana informacja . . . . .	13
2.4	Język maszyn . . . . .	14
2.5	Konwencje przekazywania informacji — w stronę języka „dla ludzi” . . . . .	15
2.5.1	Poczta elektroniczna . . . . .	16
2.5.2	Grupy dyskusyjne . . . . .	16
2.5.3	Transfer plików . . . . .	16
2.5.4	Katalogi informacji . . . . .	17
2.5.5	<i>World Wide Web</i> . . . . .	17
2.6	<i>World Wide Web Consortium</i> . . . . .	23
2.7	Przyszłość . . . . .	24
<b>3</b>	<b><i>Semantic Web</i></b>	<b>24</b>
3.1	Semantyka bez podmiotu poznającego . . . . .	25
3.2	Architektura <i>Semantic Web</i> . . . . .	27
3.2.1	Wieża <i>Sieci Semantycznej</i> . . . . .	27
3.2.2	Środowisko . . . . .	29
3.2.3	Zasoby . . . . .	30
3.2.4	Ontologie . . . . .	30
3.2.5	<i>Agenci</i> . . . . .	31
3.3	Środki formalne . . . . .	33
3.3.1	<i>Extensible Markup Language</i> . . . . .	33
3.3.2	<i>Uniform Resource Identifier</i> . . . . .	38
3.3.3	<i>Resource Description Framework</i> . . . . .	38
3.3.4	<i>Web Ontology Language</i> . . . . .	40
3.3.5	<i>Web Services</i> . . . . .	41
3.4	Przykład zastosowania . . . . .	41
<b>4</b>	<b>Problemy „semantyki stosowanej”</b>	<b>43</b>
4.1	Która semantyka? . . . . .	43
4.2	Informacja, a wiedza . . . . .	44

<b>5</b>	<b>Program badawczy</b>	<b>45</b>
5.1	Kontekst filozoficzny . . . . .	45
5.2	Semantyka i lingwistyka . . . . .	46
5.2.1	Konteksty intensjonalne . . . . .	46
5.2.2	Idealistyczno–sematyczne, czy też pragmatystyczne ujęcie zagadnienia znaczenia? . . . . .	46
5.2.3	Związki sematyczno–ontyczne . . . . .	46
5.2.4	Problem wieloznaczności . . . . .	46
5.2.5	Kontekstualizm i holizm semantyczny . . . . .	46
5.2.6	Ujmowanie znaczeń za pomocą struktur czysto syntaktycznych oraz ich przekształceń . . . . .	47
5.2.7	Lingwistyka kognitywna . . . . .	47
5.3	„Z punktu widzenia logiki” . . . . .	47
5.3.1	Aksjomaty . . . . .	48
5.3.2	<i>Access Limited Logic</i> . . . . .	48
5.3.3	O RDF i OWL . . . . .	48
5.3.4	Logiki wyższych rzędów . . . . .	50
5.3.5	Indukcja, prymitywna rekursja i generalizacja na nieskończenie wiele przypadków . . . . .	50
5.3.6	Logiki parakonsystentne . . . . .	51
5.3.7	Rozstrzygalność i problemy nierozstrzygalne . . . . .	51
5.3.8	Ujęcie metajęzyka w ramach języka przedmiotowego? . . . . .	51
5.3.9	Ujęcie elementów negatywnych . . . . .	51
5.3.10	Czy <i>Semantic Web</i> to nowa sylogistyka? . . . . .	52
5.4	Współczesny „język idealny”? . . . . .	52
5.5	Psychologia, antropologia . . . . .	53
5.6	Filozofia umysłu, neurobiologia i ewolucjonizm . . . . .	53
5.6.1	Semantic Networks . . . . .	53
5.6.2	Sieci neuronowe . . . . .	54
5.7	Systemy eksperckie . . . . .	54
5.8	Sztuczna inteligencja . . . . .	54
5.9	Wykładanie koncepcji . . . . .	55
5.9.1	Programiści i specjaliści IT . . . . .	55
5.9.2	Twórcy treści . . . . .	56
5.9.3	Popularyzatorzy . . . . .	56
5.10	Oprogramowanie . . . . .	56
<b>6</b>	<b>Zakończenie</b>	<b>56</b>

### Streszczenie

Dokument opisuje projektowaną w ramach sieci internet strukturę *Sieci Semantycznej*, która ma stanowić rozszerzenie obecnej *World Wide Web*. Projekt został omówiony w aspektach różnych dziedzin wiedzy z których bezpośrednio czerpie, lub z którymi pośrednio się wiąże. W szczególności podjęto próbę ukazania związków z niektórymi tradycyjnymi dziedzinami filozofii, a szerzej z nurtem nauk kognitywnych.

A sort of universal language or script, but infinitely different from all those projected hitherto, for the symbols and even words in it would direct the reason, and errors, except for those of fact, would be mere mistakes in calculation. It would be very difficult to form or invent this language or characteristic, but very easy to understand it without any dictionaries.

Gottfried Wilhelm Leibniz, *De Arte Combinatoria*

Every individual is at once the beneficiary and the victim of the linguistic tradition into which he has been born — the beneficiary inasmuch as language gives access to the accumulated records of other people's experience, the victim in so far as it confirms him in the belief that reduced awareness is the only awareness and as it bedevils his sense of reality, so that he is all too apt to take his concepts for data, his words for actual things.

Aldous Huxley, *The Doors of Perception*

# 1 Wstęp

Zestawienie słów „zastosowanie” i „semantyka” użyte w tytule niniejszej pracy może budzić wątpliwości. Sieci informatyczne, jak sama nazwa sugeruje, służą przekazywaniu informacji, te zaś zawsze wyrażone są w określonym — choć szeroko pojętym — języku. Wspomniane wątpliwości w duchu Wittgensteina mógłbym wyrazić w następujących zdaniach: „W jaki sposób zastosować w stosunku do języka *to*, co *uwidacznia się* samo w jego zdaniach?”, lub też: „W jaki sposób zastosować w stosunku do języka *to*, co *ujawnia się w sposobie jego używania*?”<sup>1</sup>. Ten problem, choć w różnych odsłonach, pojawia się zarówno wtedy, gdy *semantykę* rozumie się jako konkretny system znaczeń przypisany strukturalnym syntaktycznym w określonych kontekstach (np. systemy logik formalnych, komputerowe języki programowania), jak i wtedy, gdy traktuje się ją jako ogólną dziedzinę wiedzy badającą relacje między znakami a obiektami do których się one odnoszą [Par99]. Postaram się ustosunkować do tych zagadnień w dalszych rozdziałach niniejszej pracy, mogę jednakowoż zasygnalizować już teraz kwestię zasadniczą. W ramach klasycznej semantyki funkcjonuje przyjmowane bezwiednie założenie, że:

przeżywanie sensu wypowiedzi językowych jest udziałem ludzkiego umysłu.

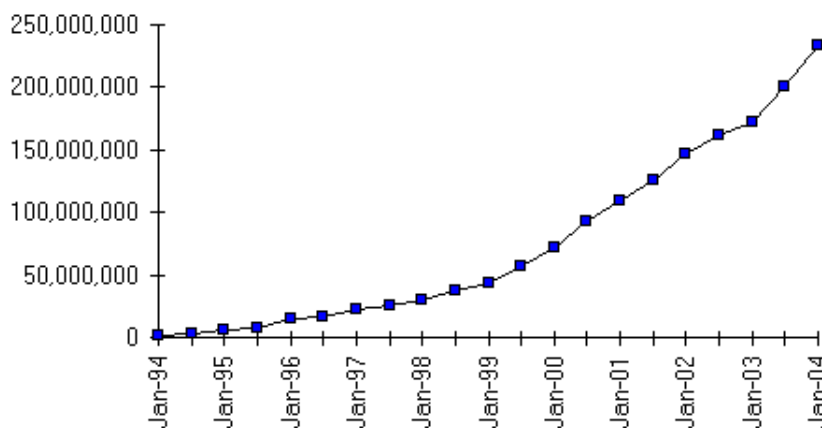
Nawet jeśli „maszyna” używa języka jako rodzaju „interfejsu komunikacyjnego” to „człowiek z chińskiego pokoju” Searle’a nie przeżywa znaczeń [Sea99]. Nawet jeśli język programowania posiada swoją *semantykę*, to reguły interpretacji sensowności konstrukcji syntaktycznych też są zaprogramowane przez człowieka. Chciałbym pokazać jakich wysiłków próbuje się dokonać obecnie, by sformułowane wyżej kryterium nie musiało być spełnione, a by dalej można było mówić o semantyce i to bez koniecznej akceptacji postulatu tzw. *silnej sztucznej inteligencji*<sup>2</sup>.

<sup>1</sup>Pierwsze sformułowanie nawiązywać by miało do *Traktatu* [Wit00b, teza 4.022], drugie natomiast do *Dociekań Filozoficznych* [Wit00a, §10]. Obydwa zaś odwołują się do poczynionej w tych dziełach próby objaśnienia takich terminów jak *znaczenie* i *sens*. Już w *Traktacie* pisze Wittgenstein o użyciu jako kryterium sensu [Wit00b, tezy 3.326, 3.327, 3.328] (choć chodzi o logiczno-syntaktyczne zastosowanie znaku, a nie „zadziałanie” językiem). Różnica między tak zwanym „pierwszym” i „drugim” Wittgensteinem jest niewątpliwa, nie należy jej jednak przeceniać. Sam fakt istnienia zarówno idealistycznego jak i pragmatystycznego nurtu w ramach semantyki wskazuje co najmniej na to, że dziedzinę tę można traktować aspektowo [Kry, Par99, HW99, Bac99]. Dostrzegam pewien związek sensu sformułowanych tu w duchu Wittgensteina zdań i mam nadzieję, że praca moja rodzaj tego związku w minimalnym stopniu naświetli.

<sup>2</sup>*Nota bene* ta nomenklatura odnośnie *Artificial Intelligence* jest stopniowo wypierana w związku z przeformulowaniem problemu w ramach teorii *Computational Intelligence*. Więcej na temat AI napiszę w rozdziale 5.8 na stronie 54.



Współczesne sieci informatyczne da się bez wielkiej straty dla nośności sformułowania sprowadzić do jednego terminu — **internet**<sup>3</sup>. Ta sieć o światowym zasięgu a równocześnie wciąż imponującym postępie przyrostu (rysunek 1) nie bez przyczyny porównywana jest do pajęczyny czy połączeń



Rysunek 1: Ilość węzłów–domen podłączonych do internetu w latach 1994–2004  
(źródło: [ISC])

neuronowych. Najbardziej adekwatnym matematycznym narzędziem opisu takiej struktury jest teoria grafów<sup>4</sup>. Dla dalszych rozważań nie jest istotne ścisłe związanie ich z internetem, lecz raczej z pewnymi możliwościami, które on udostępnia jako sieć przekazywania informacji. Jednakowoż internet istnieje i to właśnie na jego bazie powstają opisywane dalej koncepcje. Jedna z nich, szczególnie warta filozoficznej refleksji, nosi miano *Sieci Semantycznej*<sup>5</sup> i to ona właśnie dała mi asumpt do napisania tej pracy<sup>6</sup>.

## 2 Internet

Nie będę próbował w sposób ścisły zdefiniować, czym jest internet. W większości wypadków, gdy używa się tego terminu, odnoszony on jest do pewnych wybranych aspektów zjawiska sieci informatycznych, w szczególności zaś eksplikuje się go w kontekście codziennego użytkowania „zglobalizowanych” komputerów. Doświadczenie internetu, które jest udziałem przeciętnej jego użytkowniczki,

<sup>3</sup>Za pomocą czcionki pogrubionej chciałbym wyróżnić specyficzne terminy wprowadzane do tekstu po raz pierwszy.

<sup>4</sup>W pracy wykorzystane zostaną pojęcia z zakresu teorii grafów takie jak: *węzeł*, *krawędź*, *ścieżka*, *cykl*, *korzeń*, *drzewo*.

<sup>5</sup>Stosowna terminologia dotycząca omawianych zagadnień nie została jeszcze utrwalona w publikacjach polskojęzycznych. Pozwalam sobie zatem na podanie własnych tłumaczeń niektórych terminów. Tak uczyniłem w tym miejscu z określeniem *Semantic Web*. Jako jego synonim będę używał czasem także skróconego terminu *sieć*. Z kolei zapis „sieć” odnosić się będzie do internetu jako takiego. Staralem się równocześnie w całej pracy nie przekładać angielskich skrótów i terminów technicznych dla jak najwierniejszego zachowania oryginalnego sensu oraz stosownych kontekstów odniesień.

<sup>6</sup>Sam projekt *Semantic Web* jest opisany w rozdziale 3 na stronie 24.

nie rodzi krytycznego stosunku do rzeczy samej, podobnie jak nie trzeba być zegarmistrzem, by odnieść wszystkie korzyści z faktu posiadania zegarka. Idąc dalej za tą analogią, nie mam zamiaru opisywać w tym miejscu trybików „mechanizmu zegarowego”. Chciałbym skupić się raczej na perspektywach, które daje fakt posiadania chronometru, oraz jeszcze wcześniej — fakt odkrycia przez ludzkość możliwości pomiaru czasu.

Niezależnie od oceny stopnia postępu (czy też regresu) w różnych dziedzinach ludzkiej wiedzy, nikt chyba nie będzie skłonny zaprzeczyć, że technologia, a zatem i internet, jej sztandarowy wykwit, rozwija się. Towarzyszą temu coraz bardziej wyszukane sposoby organizowania materii służące fizycznej reprezentacji informacji, tak by umożliwić jej przekazywanie. Moje rozważania chciałbym celowo, o ile jest to oczywiście możliwe, podjąć poza paradygmatem fizykalno–technologicznym. Jeśliby odwołać się do specyficznego profesjolektu, można powiedzieć, że próbuję dokonać zabiegu oddzielenia wszystkiego co „*hardware’owe*” od tego co „*softwarowe’owe*”, tego co „sprzętowe” od tego co należy do „oprogramowania”. Dostrzegany w praktyce codziennej użytkowników sieci wzrost stopnia niezawodności, dostępności, szybkości etc. stanowi powiększenie się puli specyficznych możliwości. Sposób, w jaki mogą one zostać wykorzystane, przynależy do innego, niefizykalnego porządku dociekań.

Nie trzeba być neurologiem by przeżywać akty świadomości i nie trzeba też być konstruktorem procesorów, by być producentem przetwarzanej przez nie informacji. Analogia ta kuleje o tyle, że w drugim przypadku analogon tak dzisiaj poszukiwanego **engramu** możemy z grubsza uchwycić<sup>7</sup>. Nie zmienia to jednak faktu, że ze względów (o ironio) pragmatycznych zakładamy, że taki byt jak *informacja* ma charakter *wirtualny* i jest w istotny sposób różny od swojej fizycznej implementacji.

## 2.1 Zakres znaczeniowy pojęcia

Słowo *internet* zagościło w naszym języku z chwilą, gdy przestało być terminem używanym jedynie przez fachowców zajmujących się komputerami oraz wszystkim co w orbicie tego tematu się znajduje. Słowo to trafiło z bardzo specyficznego profesjolektu do powszechnych słowników kultur technologicznie rozwiniętych, dzieląc w tym względzie los określeń innych wynalazków o masowym charakterze, takich jak radio, telefon czy telewizor. Zważywszy jednak na to, że termin „internet”

<sup>7</sup>Poszukiwania tak zwanego „ślądu pamięciowego” zmierzają nie tylko do odkrycia konkretnych fragmentów masy mózgu (prawdopodobnie kory mózgowej), które przechowywałyby konkretne dane — analogicznie jak domeny magnetyczne czy kondensatory o odświeżanym ładunku w przypadku pamięci komputera — lecz przede wszystkim do odkrycia sposobu w jaki zaimplementowana jest „komórkowa i molekularna baza pamięci” oraz na czym *de facto* od strony komputacyjnej polega zapamiętywanie [Gal90, Bad99, Squ99]. Oliver Sacks opisuje pacjentów, którzy w wyniku bardzo określonych urazów mózgu stracili, lub też „przywołali” pamięć bardzo określonych rzeczy (obrazów, melodii). „Ostateczna postać reprezentacji mózgowej musi być «sztuką» — artystycznym krajobrazem i melodią przeżywania i działania”. Już dziś możliwe jest, jak pokazał to Penfield, bardzo konkretne neurochirurgiczne, czy też może raczej „psychochirurgiczne” ingerowanie w ten krajobraz — niestety polegające jedynie na destrukcji — wywołujące „bardzo szczególną amnezję” [Sac96]. Główny bohater filmu  $\pi$ , genialny matematyk, nie mogąc znieść swej wciąż aktywnej „świadomości obliczeniowej”, korzystając z mapy kory mózgowej niszczy własny „nadaktywny” ośrodek komputacyjny. Nie używa jednak do tego celu skalpela, lecz... wiertarki.

stosowany jest w dwóch różnych kontekstach znaczeniowych, należy dokonać w tym miejscu paru uściśleń.

Pierwsza płaszczyzna odniesienia związana jest z bazą technologiczną dostarczającą środków przekazywania informacji na szereg precyzyjnie zdefiniowanych sposobów, które jednakowoż muszą znaleźć ostatecznie pewien „wspólny mianownik”. Mam tu na myśli fizyczną implementację *internetu* — przewody elektryczne lub światłowody, urządzenia które są z ich pomocą łączone, oraz tak zwane **protokoły sieciowe** które pozwalają tym urządzeniom nadawać **sygnały**, a ich analogom po drugiej stronie przewodu odbierać je i w sposób jednoznaczny interpretować. Jednym słowem: to wszystko, o czym większość ludzi używających na codzień dobrodziejstw **poczty elektronicznej**<sup>8</sup>, czy stron **WWW**<sup>9</sup> nie tylko nie wie, lecz wręcz wiedzieć sobie nie życzy (z powodów całkowicie zresztą dla mnie zrozumiałych). Co ciekawe, użytkowniczkę „sieci” właśnie w ten sposób ją zazwyczaj konceptualizują. Przyjmują, że stanowi ona po prostu gotową do użytku — empirycznie dostępną syntezę elementów przed empirią ukrytych z racji chociażby ich energetycznego, ulotnego charakteru. Traktują ją jako środek komunikacji zachowujący się w pożądanym sposób, zakładając równocześnie, że „internet” to nic innego jak miliardy mknących równocześnie przewodami elektronów. Ukute przez Michała Hellera określenie „cywilizacja czarnej skrzynki” [Hel00] jest tu jak najbardziej adekwatne. Interesuje nas tylko i wyłącznie sposób zachowania pewnych automatów, z którymi przychodzi nam obcować. Nie chcemy wiedzieć co kryją ich wnętrza. Takie powszechne „behawioralne” podejście może jednakowoż przesłonić wydawało by się najistotniejszy aspekt całego zagadnienia. Kartezjusz arbitralnie odebrał zwierzętom prawo posiadania wszelkich procesów świadomościowych, postrzegając je jako bardzo skomplikowane mechanizmy. Na mechanizm trudno jest spojrzeć inaczej niż przez pryzmat sposobów w jakich się nam zewnętrznie prezentuje, mając równocześnie w domyśle całą jego ukrytą złożoność. Podobnie rzecz się ma z internetem. Fakt, że istotą jego funkcjonowania jest wymiana treści, oraz umożliwienie użytkownikom przeżywania znaczeń, pozostaje w dużym stopniu dla tych ostatnich przezroczysty<sup>10</sup>.

Ta druga, nie technologiczna, lecz „semantyczna” płaszczyzna odniesienia związana jest z całością informacji, która potencjalnie może zostać w ramach infrastruktury technologicznej przekazana. Oczywiście nie chodzi o to, żeby dążyć do zrealizowania za wszelką cenę tej potencjalności, lecz o to, że dla użytkowniczkę internetu obcowanie z pewnymi „zdeponowanymi” w nim informacjami uchodzi za szczególną wartość. Analogia z biblioteką jest tu bardzo czytelna. Nikt przy zdrowych zmysłach nie będzie starał się wypisać rewersów na wszystkie książki, choć jest to teoretycznie możliwe, lecz zamówi tylko te określone według własnych preferencji. Właśnie w owej możliwości tkwi siła koncepcji leżącej u podwalin obydwu instytucji. Gdyby było inaczej, każdej wystarczyłby całko-

<sup>8</sup>Patrz w rozdziale 2.5.1 na stronie 16.

<sup>9</sup>Patrz w rozdziale 2.5.5 na stronie 17.

<sup>10</sup>Nie chcę oczywiście by z tej analogii wyciągnięty został wniosek o samoświadomości internetu, choć i takich pomysłów nie brakuje, by przywołać chociażby te, które dopatrują się w „sieci” realizacji *Noosfery* Teilharda de Chardin [Pri].

wicie własny księgozbiór domowy (podobnie najczęściej występująca, a często też jedyną książką w domach wielu Polaków jest książka kucharska, co skłania do refleksji, że potrzeba dostępu do informacji jest ostatnią w hierarchii potrzeb, daleko za zaspokojeniem roszczeń żołądka, libido i morfeusza).

Podsumowując: chciałbym zająć się tutaj internetem w jego płaszczyźnie semantycznej. Potraktować go jak bibliotekę, z której mogę wypożyczyć określoną książkę, a książki tej nie chcę postrzegać jako zbioru paru tysięcy znaków drukarskich, zapewniających parę setek kartek, lecz jako dostępną mojej świadomości wiązkę pewnych treści. W dalszej części pracy chciałbym pokazać ponadto, na jakich warunkach wiedza taka może zostać skodyfikowana, tak by zawarta ona była w samej informacji, niezależnie od próbującego ją przyswoić ludzkiego umysłu.

## 2.2 Geneza

Światowa sieć komputerowa jest tworem, który podlega procesowi ciągłego rozwoju. Informacja nie będąc w konieczny sposób związana z pewną fizyczną reprezentacją (implementacją) poddawana jest ciągłym przekształceniom. Tej wolności nie posiadają inne tradycyjne formy przekazu informacji, takie jak pismo, czy nagranie dźwiękowe. W związku z tym, że wszystkie atomowe zmiany struktury sieci są częstokroć ze sobą powiązane, mówić tu możemy o czymś na kształt ewolucji biologicznej, a nie o realizacji ustalonego projektu naukowego. W toku tej ewolucji wyłoniło się bardzo wiele koncepcji co do sposobów reprezentacji i przekazywania informacji, z których wiele nie wytrzymało jednakowoż próby czasu. Spróbuję pokrótce opisać tę ewolucję.

Protoplastą internetu jest telegraf. Funkcja telegrafu polega na przesyłaniu komunikatów. Wykorzystuje się do tego celu urządzenia nadawcze i odbiorcze, oraz pewien kod komunikacyjny (np. alfabet Morse'a). Proste przekazywanie informacji z jednego miejsca w inne nie nastręcza trudności. Gorzej gdy należy do takiego przekazania użyć wielu telegrafów, często różnych typów (świetlny, elektryczny, radiowy). Załóżmy, że znaleźliśmy się w realiach początku wieku XX. W każdym większym mieście jest telegraf, ale chcąc przesłać naszą wiadomość z Warszawy do Budapesztu musimy zdecydować w jaki sposób ona tam dotrze. Możemy zatelegrafować najpierw albo do Krakowa albo do Wrocławia, ufając że obsługa któregoś z tamtejszych telegrafów będzie na tyle sumienna by nasz komunikat skierować dalej odpowiednio do Koszyc lub do Pragi. Co będzie się z nim działo później pozostaje już dla nas tajemnicą.

Telegraf został skutecznie zastąpiony przez wynalazek radia, choć nadawanie na odległość sygnałów radiowych bazowało na samym początku właśnie na przekazywaniu informacji w sposób zakodowany (alfabet Morse'a), a nie bezpośrednio głosu, jak ma to najczęściej miejsce dzisiaj<sup>11</sup>. Jednakowoż koncepcja pozostała, a powstanie technicznych możliwości zautomatyzowania i przyspieszenia

<sup>11</sup>Paradoksalnie dzisiaj również przekazywanie głosu ma charakter cyfrowy — podlega na kodowaniu (digitalizacji) i kompresji danych dźwiękowych, co często jest połączone z ich szyfrowaniem. Dotyczy to w szczególności telefonii cyfrowej i komórkowej, a w przyszłości w coraz większym stopniu także internetowej technologii *Voice over IP* — przekazywania głosu za pomocą internetu.

tradycyjnego procesu komunikowania przez wyeliminowanie człowieka jako elementu pośredniczącego zaowocowało załóżkiem sieci znanej nam dzisiaj jako internet.

Cały koncept miał militarną proveniencję<sup>12</sup> — usiłowano stworzyć sieć przekaźników informacji działającą całkowicie automatycznie, a przy tym zdecentralizowaną, to znaczy nie posiadającą miejsc szczególnie wyróżnionych. Wszystko po to, by zniszczenie fragmentu sieci (w końcu były to czasy „zimnej wojny”) nie spowodowało załamania całej komunikacji. Informacja powinna sama poszukiwać optymalnej ścieżki do swojego punktu przeznaczenia. Jak się z czasem okazało, „anarchistyczna” w swej istocie koncepcja nie może zostać skutecznie wykorzystana w totalizującym zamyśle, jakim jest doktryna wojenna.

W roku 1969 powstaje ARPANet<sup>13</sup>, i choć na samym początku węzłami sieci są instytucje militarne i pracujące na ich zamówienie uczelnie, bardzo szybko jednak to te drugie zaczynają dominować w wykorzystywaniu nowej technologii. Trudno się temu dziwić; to właśnie ośrodki akademickie znane są z produkowania wielkich ilości informacji, którymi chciałyby równocześnie wymieniać się z innymi podobnymi sobie instytucjami. ARPANet przekształca się z czasem w internet, konsolidując różne inne powstałe w międzyczasie sieci.

Internet to miliony połączonych ze sobą węzłów. Każdy z nich może odebrać, wysłać, czy też przekazać dalej pewien strumień informacji. Także typowy domowy komputer może zostać „podłączony” jako tego rodzaju węzeł. Nagła „eksplozja” internetu, która nastąpiła w latach 90., związana jest niewątpliwie z wynalezieniem i powszechnym zastosowaniem pomysłu *hipertekstu*, który ukonstytuował to, co szumnie nazwano *Siecią Światowej Skali* czyli *World Wide Web*<sup>14</sup>.

Z dzisiejszej perspektywy internet nie jawi się już wyłącznie jako odległa nowinka techniczna. Zaczyna być medium o poziomie wykorzystania analogicznym do telefonu, telewizji czy radia. Choć ilość osób posiadających telefony komórkowe wzrasta szybciej niż ilość użytkowników „sieci”, to właśnie ta ostatnia stanowi „rewolucję” o tyle, że proponuje zupełnie nowy paradygmat komunikowania się.

## 2.3 Warstwa transportowa, a przekazywana informacja

W rozdziale 2.1 na stronie 10 pisałem o dwóch różnych kontekstach znaczeniowych, w których pojawia się termin „internet”. Zadeklarowałem również, że mam zamiar zajmować się tym bardziej abstrakcyjnym poziomem, rzetelność wymaga jednak, by pokazać wzajemne związki obu kontekstów, co na szczęście można zrobić również w sposób abstrakcyjny, nie wdając się dzięki temu niepotrzebnie w szczegóły techniczne.

Problem fizycznego sprzętu służącego transmisji danych jedynie tutaj zasygnalizuję. Ponad nim

<sup>12</sup>Bardzo szeroko opisaną historię internetu znaleźć można w [LCC<sup>+</sup>].

<sup>13</sup>ARPA to skrót od *Advanced Research Projects Agency*, będącej częścią Departamentu Obrony USA.

<sup>14</sup>Nad powstaniem WWW warto zatrzymać się nieco dłużej, co uczynię w rozdziale 2.5.5 na stronie 17.

znajduje się tak zwana *warstwa transportowa* internetu. Przytoczona już analogia z telegrafem<sup>15</sup> także w tym miejscu okaże mi się pomocna. *Warstwa transportowa* to protokoły sieciowe, czyli stale dopracowywane „alfabety Morse’a” dla maszyn — cały zestaw „dialektów” określających standardy kodowania informacji po stronie nadawcy i dekodowania po stronie odbiorcy, tak by skutecznie je przekazywać<sup>16</sup>.

Następny poziom stanowi sama informacja. Musi być ona całkowicie „przetwarzalna” do postaci cyfrowej (digitalizacja), ale co jest tu niezmiernie ważne, nie można jej z tą postacią utożsamić. Ze względów o których pisałem na początku niniejszego rozdziału, dokonać należy przesunięcia punktu ciężkości z fizycznej reprezentacji (implementacji) komunikatu na treści z nim związane, niezależnie od tego czy kierowane są one do człowieka, czy też do maszyny.

Dla abstrakcyjnej komunikacji maszyn wystarczają zatem:

- fizyczna postać sygnału (reakcje chemiczne, elektryczne, mechaniczne, świetlne).
- jej logiczna interpretacja, czyli usystematyzowanie sygnałów w kody poddające się jednolitym kryteriom przekazywania i weryfikacji.

Natomiast aby komunikację maszyn w jakiś sposób wykorzystać potrzebna jest także:

- struktura pewnego języka, operującego jednostkami leksykalnymi możliwymi do zakodowania w poprzedniej warstwie, czyli bazująca na usystematyzowaniu opisanego wyżej odwzorowania tego, co algebraizowalne, w to, co jest czysto fizyczne.

## 2.4 Język maszyn

Podstawowy kod internetu i informatyki w ogóle składa się ze ściśle określonego zbioru znaków. Nie są to znaki reprezentowane w postaci symboli graficznych, są one natomiast jednoznacznie interpretowane przez maszyny stosownie do określonych reguł. Zbiór ten posiada 256 elementów ( $2^8$  — 8 bitów (cyfr) w systemie dwójkowym), przy czym ich kolejność jest istotna — tworzą one ciąg. Dzięki temu możliwe jest dokonanie zabiegu enumeracji dla każdego ze znaków stosowanych w danym kontekście komunikacyjnym. Np. można przyporządkować numery określonym znakom alfabetu łacińskiego. Komunikacja przebiega prawidłowo pod warunkiem, że interpretacja „numerów” jest taka sama zarówno dla systemu nadającego, jak i odbierającego cyfrowe treści.

Każdy rodzaj przesyłanych za pomocą internetu informacji składa się z dowolnej długości sekwencji opisanych wyżej znaków. W ramach każdej takiej sekwencji, czyli komunikatu, wydzielić można podsekwencje stanowiące jednostki leksykalne. Analogicznie odnajdujemy słowa w ciągu

<sup>15</sup>Patrz w rozdziale 2.2 na stronie 12.

<sup>16</sup>W większości współczesnych protokołów sieciowych dodatkowym ważnym elementem jest możliwość korekcji błędów powstałych w czasie transmisji.

czytanego tekstu dzięki rozdzielającym je odstępom. Widać tutaj, że podobnie jak w przypadku alfabetu Morse’a zakodować możemy dowolne zdania. Niestety moc eksplanacyjna „analogii telegraficznej” zaczyna się w tym miejscu stopniowo wyczerpywać. Komunikat wcale nie musi być skierowany do człowieka, w którym to przypadku maszyny stanowią tylko elementy biernie przekazujące. Z równym powodzeniem one same mogą interpretować opisane struktury leksykalne. We współczesnym internecie mamy doczynienia z rozwiązaniami hybrydowymi, przy czym bardzo wyraźna jest tendencja, by komputery przeprowadzały coraz większą część analizy transmitowanych danych. Pojawia się zatem zagadnienie wykorzystywania języka oraz metajęzyka.

Przyjmijmy dla uproszczenia, że poziom leksykalny przesyłanych danych stanowią słowa języka angielskiego, kodowane za pomocą alfabetu łacińskiego, który z kolei poddawany jest enumeracji w sposób opisany wyżej. Wyobraźmy sobie że zakodowany jest dokument opisujący zagadnienie *Sieci Semantycznej*. W tekście występować będzie sekwencja znaków „*Semantic Web*”. Kryteria według których czytelniczka byłaby skłonna przypisać temu fragmentowi znaczenie tytułu są dość mętne. Jeśli nie dysponujemy żadnym rozróżnieniem krojów czcionek drukarskich, czy też ich wielkości — a tak jest w przypadku nienacechowanego, cyfrowego „strumienia” danych — to pozostaje jej na przykład zbadanie niestandardowych odstępów wokół tego fragmentu tekstu — kryterium zdecydowanie nie formalne. Jeśli chcemy, by komputer był w stanie jednoznacznie interpretować wzminkowany fragment musimy dostarczyć „maszynie” kryteria ściśle. Ich specyfikacja może być wpisana w dane *implicite* lub *explicite*. W pierwszym przypadku sama struktura — pozycja elementu licząc od początku strumienia — decyduje o sposobie interpretacji danych. W drugim dokonuje się zabiegu zarezerwowania pewnych znaków lub całych struktur leksykalnych jako meta–oznaczeń wyrażonych w ramach języka przedmiotowego. W rozważanym przypadku mógłby to być język określający strukturę dokumentu z rozróżnieniem bloków składowych takich jak tytuł i na przykład akapit. Najczęściej stosowana dziś metoda polega na zarezerwowaniu pewnych znaków, lub całych jednostek leksykalnych tak, by kodowały one terminy metajęzykowe, a nie język przedmiotowy.

## 2.5 Konwencje przekazywania informacji — w stronę języka „dla ludzi”

Z poprzedniego rozdziału wynika, że istnieje możliwość stworzenia nieskończenie wielu meta–języków służących przekazywaniu informacji w ramach internetu, i faktycznie bardzo wiele takich języków powstało, lecz silną pozycję zyskały tylko niektóre z nich. Spełniają one dziś funkcję powszechnie akceptowanych standardów<sup>17</sup>. Jako że koncepcja *Sieci Semantycznej* również wiąże się z wprowadzeniem pewnego nowego języka, przedstawię pokrótce różne używane w ramach internetu konwencje przekazywania informacji, z których *Semantic Web* czerpie. Kolejność, w której zostaną one opisane,

<sup>17</sup>Niektóre specyfikacje opisane są w tak zwanych dokumentach RFC, co stanowi skrót od *Request for Comments* i pokazuje otwarty (*open, free, public domain*) — związany ze swoistym „konsensusem” charakter wielu internetowych standardów. Wraz z szeroką obecnością w ramach internetu instytucji komercyjnych ta tendencja ciąży w stronę tzw. rozwiązań „zamkniętych” (*closed, proprietary*), czemu towarzyszą zupełnie zrozumiałe protesty bardziej świadomych użytkowników *sieci*, której to rozwój bez owej „otwartości” może zostać skutecznie zahamowany.

nie jest przypadkowa. Poza kryterium chronologicznym istotne jest też związane z nim kryterium stopnia „autonomizacji informacji”. Chodzi o fazy, gdzie zaczyna ona coraz bardziej funkcjonować w odłączeniu od pojedynczego ludzkiego umysłu, który ją „absorbuje”.

### 2.5.1 Poczta elektroniczna

„*Electronic Mail*” nazwę swoją zawdzięcza temu, że zasada jej działania jest zupełnie analogiczna do zasady działania tradycyjnej poczty. Nadawca może sporządzić wiadomość zawierającą pewną ilość tekstu i zaadresować ją do dowolnej ilości odbiorców. Cała instytucja współpracujących urzędów pocztowych zostaje zastąpiona *warstwą transportową* internetu. Rzeczą wartą zaznaczenia jest niespotykany w dotychczasowych środkach komunikacji nacisk położony na wartość samej przekazywanej treści. Możliwość zacytowania dowolnego fragmentu listu korespondenta i ustosunkowania się właśnie do niego jest czymś zasadniczo odbiegającym od reguł tradycyjnej sztuki epistolarnej. Nowe możliwości, które dzięki temu powstają, zostały skrzętnie wykorzystane i skodyfikowane w formie (wtórnie spisanych) konwenansów zwanych „netykietą”, z wielką korzyścią dla precyzji, klarowności, a co za tym często idzie również siły i poziomu merytorycznego argumentów formułowanych w ramach „dyskusji” prowadzonych za pomocą wymiany e-maili. Konwencje wskazujące jak oznaczyć adres (słynny znak „@”), wyświetlić tytuł, jak pokazać adresata lub zaznaczyć cytaty i jak się podpisać, stanowią fenomen o tyle, że pozwalają prześledzić typ genezy meta-języków komputerowych.

Pewną mutacją koncepcji poczty elektronicznej są popularne dzisiaj aplikacje komunikacyjne (*messenger*). Ich odmienność od tej pierwszej polega na tym, że przekazana wiadomość (najczęściej bardzo krótka) powinna zostać zaprezentowana odbiorcy jak najszybciej, by umożliwić mu też jak najszybszą odpowiedź. W tym sensie programy tego typu stanowią internetową wersję telefonu, gdzie zamiast fali akustycznej, nośnikiem informacji jest zdigitalizowany tekst.

### 2.5.2 Grupy dyskusyjne

Grupy dyskusyjne sprawiają, że poczta elektroniczna zyskuje globalny charakter. Wiadomość wysłana przez jednego internautę może zostać udostępniona na żądanie dowolnej liczbie innych użytkowników „sieci”. Wysyłamy e-mail, adresując go nie do konkretnego odbiorcy, lecz do publicznej skrzynki reprezentującej grupę osób dyskutujących na ten sam temat. W efekcie możemy uczestniczyć w tematycznych dyskusjach o globalnym zasięgu, co jest niewątpliwą zaletą, o ile do wymiany poglądów nie włączy się ktoś nie akceptujący przyjętych przez nas standardów komunikacyjnych.

### 2.5.3 Transfer plików

Zgodnie z tym co pisałem w rozdziale 2.4 na stronie 14, dane komputerowe muszą być wyrażone w postaci **binarnej**, którą da się przełożyć na ciągi liczb całkowitych o wartościach od 0 do 255. Sekwencja takich numerów może stanowić pewien zamknięty zbiór informacji nazywany **plikiem**. Np.:



ciąg 120000 takich liczb może opisać zdigitalizowaną fotografię o wymiarach 200 na 200 punktów<sup>18</sup>. W przypadku konwencji opisania tekstu, najczęściej to pojedyncza liczba całkowita z zakresu od 0 do 255 koduje jedną literę, lub znak specjalny typu *nowa linia*, *tabulator*, *odstęp*. Pliki mogą zostać zinterpretowane w odpowiedni sposób (grafika, tekst, dźwięk) przez przechowujący je komputer.

FTP to skrót oznaczający **protokół wymiany plików** (*File Transfer Protocol*). Umożliwia on zdeponowanie pewnej ilości plików w określonym „miejscu” sieci — na **serwerze** (np. na uniwersytecie Berkley w USA) i pobranie wybranych z nich w zupełnie innym (np. na jednym z komputerów — stacji roboczych na Uniwersytecie Szczecińskim). Zdefiniowane zostały reguły określające standardy kodowania przesyłanych informacji, oraz język pozwalający na jednoznaczne wskazanie konkretnych zasobów.

Z punktu widzenia dalszych rozważań najbardziej interesujące są **pliki tekstowe**, ponieważ to one mogą zawierać informacje w postaci pozwalającej na poddanie ich różnorodnym analizom leksykalnym, począwszy od ortograficznych, a skończywszy na gramatyczno–semantycznych. W dalszych rozdziałach pracy zostaną opisane możliwości stosowania właśnie tych szczególnie abstrakcyjnych poziomów analiz.

#### 2.5.4 Katalogi informacji

Kiedy posiadamy już pewną ilość plików zdeponowanych w różnych miejscach „sieci”, automatycznie pojawia się potrzeba zindeksowania ich zawartości w celach wyszukiwawczych. Stopniowo wyłoniły się dwie zasadnicze grupy tego typu **serwisów indeksujących**.

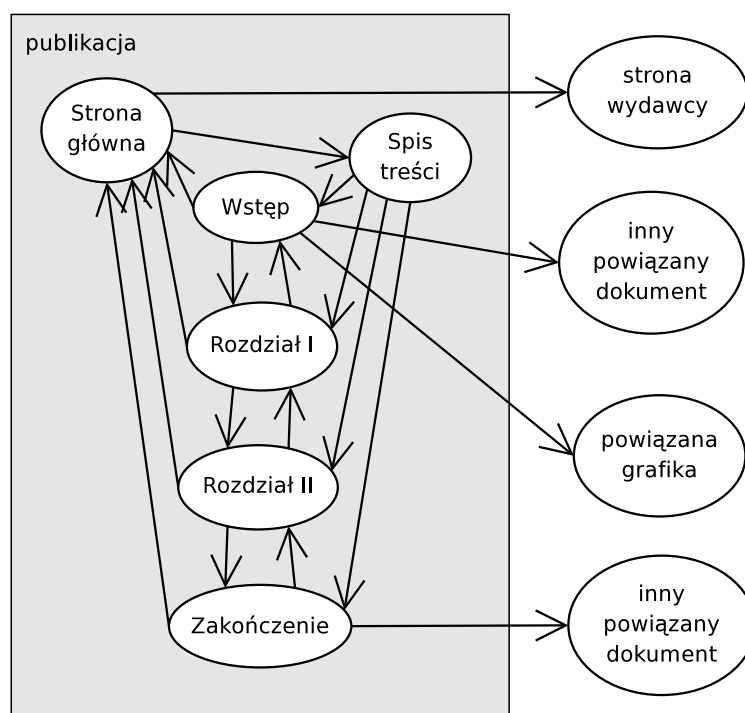
Pierwsza z nich, to **katalogi** porządkujące zbiory danych w ramach kategorii tematycznych. Analogia z katalogiem tematycznym tradycyjnej biblioteki jest bardzo wyraźna, pozycja w katalogu sieciowym wskazuje jednak na określony plik, a nie na tradycyjne wydawnictwo.

Drugą grupę stanowią automaty indeksujące zawartość tekstową wyszukanych (w idealnej sytuacji wszystkich) w obrębie sieci plików. Tu z kolei odpowiedniejsza będzie analogia z indeksem terminów kluczowych umieszczanym na końcu książki. Wpis indeksu zawiera termin, oraz strony na których on występuje. Możliwości wykorzystania takiego rozwiązania zależą od ilości przeanalizowanych zasobów. Serwis tego typu nazywany jest często **wyszukiwarką internetową** (*search engine*) i z punktu widzenia użytkownika prezentuje się w ten sposób, że wprowadzenie określonego słowa powoduje wyświetlenie listy wszystkich dokumentów w których ono występuje. Warto zaznaczyć że analogicznie zindeksować można również np. zawartość archiwów grup dyskusyjnych.

#### 2.5.5 World Wide Web

Koncepcją konstytutywną dla **World Wide Web** (WWW), czyli *Sieci Światowej Skali* jest wynalazek tak zwanego **hipertekstu** [W3Hb]. W tym miejscu najczytelniejsza będzie analogia ze słownikiem

<sup>18</sup>Potrzebne są 3 liczby na każdy punkt —  $200 \times 200 \times 3 = 120000$ . Chodzi o wartości składowych kolorów czerwonego, zielonego i niebieskiego w addytywnym modelu barw.



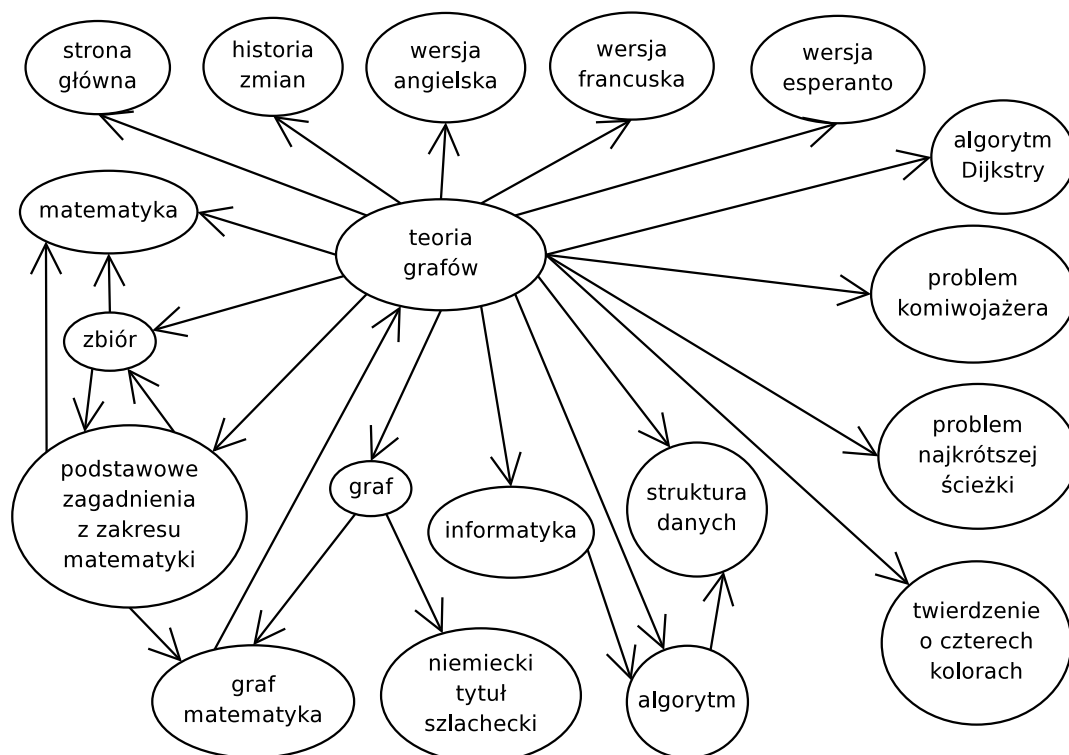
Rysunek 2: Przykład struktury publikacji HTML w reprezentacji *grafu skierowanego*

lub encyklopedią. Często zdarza się, że jedno hasło definiowane jest z wykorzystaniem innych, powiązanych, o czym dowiadujemy się na podstawie metainformacji zawartej w formułach: „zobacz także”, „porównaj”, itp. Analogicznie funkcjonują odsyłacze bibliograficzne — wskazują na pewien dokument, do którego odnosi się treść obecnego tekstu. W przypadku internetu odpada problem ponownego wyszukiwania informacji. Odsyłacz umożliwia nam bardzo łatwe przeniesienie się do wskazywanego przezeń dokumentu (pobrany zostanie plik go zawierający). Zamiast wyrażenia *explicit* kategorii „zobacz także”, czy też „porównaj”, pojawia się inna forma oznaczająca to samo — z reguły fragment tekstu wyróżniony jest podkreśleniem i innym kolorem czcionki.

Reprezentacji informacji narzuca się w tym przypadku strukturę, którą matematycznie może opisać *graf skierowany* (*directed graph*). *Węzłem* (*wierzchołkiem*) jest tutaj dokument tekstowy a *krawędź*<sup>19</sup> zaczyna się w nim, kończy zaś w innym dokumencie — *węźle*. Rysunek 2 przedstawia *graf* dla hipotetycznej publikacji, która w ramach jednego **serwisu internetowego** (*site*) zawiera „wskazujące” na siebie nawzajem **strony**, odpowiadające częściom tradycyjnego dokumentu, takim jak spis treści czy rozdziały, a ponadto odwołuje się do zasobów zewnętrznych — grafiki oraz innych tego typu plików tekstowych wskazujących potencjalnie na dalsze zasoby. Rysunek 3 na następnej stronie przedstawia *graf* dla definicji hasła „teoria grafów” [Wika] zaczerpniętej z Wikipedii — wolnej, międzynarodowej, tworzonej przez internetową społeczność (*community*) encyklopedii [Wikb].

Oczywiście oba *grafy* pokazują układ elementów i relacji między nimi wyekstrahowany z kon-

<sup>19</sup>Używany jest w tym przypadku termin *link*, czyli połączenie, lub *anchor* — kotwica. Można zatem mówić o *zakotwiczeniu*.



Rysunek 3: Dokument HTML „Teoria grafów” w reprezentacji *grafu skierowanego*

tekstu znacznie większej, a w rzeczywistej „przestrzeni internetu” wciąż rozrastającej się „sieci”. Zarówno „strona wydawcy” jak i wszelkie definicje z Wikipedii ukazane na obrzeżach diagramu z rysunku 3 powiązane są z kolejnymi, nie pokazanymi *węzłami*.

Stosunkowo prosta w swych założeniach koncepcja *hipertekstu* miała tak rewolucyjne następstwa, że ich ogrom ciężko będzie opisać i w mojej pracy bynajmniej nie aspiruję do tego by to uczynić. Eksplozja internetu związana jest z rozrostem gigantycznej **sieci asocjacyjnej**, reprezentującej paradygmat odmienny od katalogu, który ma układ hierarchiczny. Nie dziwi zatem często stosowana na określenie tego fenomenu nazwa — „światowa pajęczyna”.

Koncepcja WWW zrodziła się w roku 1990 w umyśle Tima Bernersa-Lee [BLc] pracującego wówczas w laboratoriach CERN pod Genewą<sup>20</sup>. Był to pomysł na ujednoczenie metod dostępu do różnych gromadzących informacje systemów komputerowych. Ujednoczeniu miało towarzyszyć równoczesne uproszczenie sposobu obsługi takich systemów dla ich końcowych użytkowników, którymi byli ludzie pracujący w różnych ośrodkach naukowych. Celem nadrzędnym zatem była możliwość jak najskuteczniejszej prezentacji wszelkich wyników badań. W konsekwencji zdefiniowano zarówno protokół komunikacyjny HTTP<sup>21</sup>, w ramach którego informacje mają być przekazywane,

<sup>20</sup>Właściwie sam pomysł zaczął się krystalizować dużo wcześniej. Już w roku 1980 Tim Berners-Lee sporządził dla CERN oprogramowanie wykorzystujące ideę *hipertekstu*. Z kolei sam termin *hipertekst* pojawił się w latach 60., w publikacjach Teda Nelsona [W3Hc].

<sup>21</sup>Skrót od *Hypertext Transfer Protocol* [W3Ha].

jak i język opisu dokumentów HTML<sup>22</sup>, rozszerzający tradycyjny format plików tekstowych o dodatkowe metainformacje. Właśnie owe metainformacje stanowią o sile całej koncepcji.

Zasobami danych w ramach WWW mogą być zarówno dokumenty tekstowe jak i pliki graficzne, a także dźwiękowe, lub każdego innego rodzaju dane, które poddają się digitalizacji (ilość typów plików wzrasta wraz ze wzrostem wydajności przeciętnego domowego komputera). Zasób jest identyfikowany za pomocą unikalnego, przypisanego mu adresu. Adres to ciąg znaków, z których część reprezentuje nazwę serwera, a część określony, zdeponowany na nim plik<sup>23</sup>. Odpowiedni program komputerowy popularnie zwany **przeglądarką internetową** jest w stanie przedstawić użytkownikowi wyspecyfikowane za pomocą adresu pliki na ekranie monitora. Takie możliwości udostępnia właśnie HTTP.

Jeżeli pobieranym plikiem jest dokument HTML, przeglądarka interpretuje zawarte w nim metainformacje, „próbując” wykorzystać je do stosownej prezentacji tekstu. Metainformacje dotyczą przede wszystkim sposobu podziału logicznego treści dokumentu i są dostarczane w postaci tzw. znaczników. Znaczniki mogą wskazywać na przykład:

- początek i koniec akapitu
- występowanie w tekście grafiki
- pogrubienie fragmentu tekstu
- informacje o autorze dokumentu

Znacznik, którego wprowadzenie miało najbardziej daleko idące konsekwencje określa, że część objętego nim tekstu odnosi się do innego pliku. Plik taki może być analogicznym dokumentem znajdującym się na tym samym lub innym serwerze. Właśnie dzięki temu użytkowniczka korzystająca ze strony WWW korzysta także z potężnej sieci asocjacyjnej, której elementem jest dana strona. Oto przykład poprawnego pliku HTML:

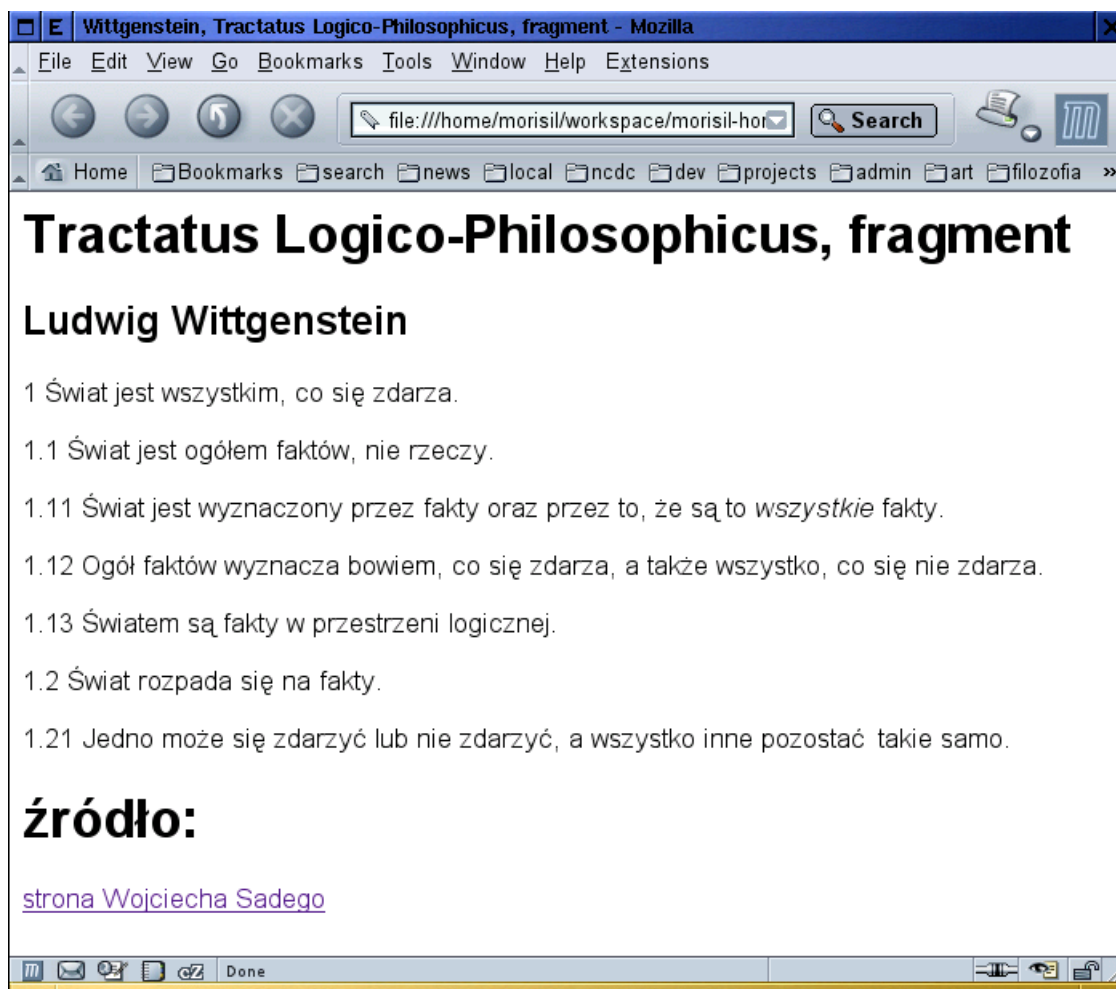
```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<html>
  <head>
    <title>Wittgenstein, Tractatus Logico-Philosophicus, fragment</title>
  </head>
  <body>
    <h2>Tractatus Logico-Philosophicus, fragment</h2>
    <h1>Ludwig Wittgenstein</h1>
    <p>1 Świat jest wszystkim, co się zdarza.</p>
    <p>1.1 Świat jest ogółem faktów, nie rzeczy.</p>
```

<sup>22</sup>Skrót od *Hypertext Markup Language* [W3Hb].

<sup>23</sup>Format takiego adresu nosi nazwę *Uniform Resource Identifier* (URI) [W3U].

```
<p>
  1.11 Świat jest wyznaczony przez fakty oraz przez to,
  że są to <i>wszystkie</i> fakty.
</p>
<p>
  1.12 Ogół faktów wyznacza bowiem, co się zdarza,
  a także wszystko, co się nie zdarza.
</p>
<p>1.13 Światem są fakty w przestrzeni logicznej.</p>
<p>1.2 Świat rozpada się na fakty.</p>
<p>
  1.21 Jedno może się zdarzyć lub nie zdarzyć,
  a wszystko inne pozostać takie samo.
</p>
<h2>źródło:</h2>
<p>
  <a href="http://sady.strona.pl/witt.tlp.fr.htm">
    strona Wojciecha Sadego
  </a>
</p>
</body>
</html>
```

Rysunek 4 na następnej stronie pokazuje w jaki sposób przedstawiona powyżej zawartość pliku HTML jest ostatecznie prezentowana w przeglądarce WWW.



Rysunek 4: Wyświetlanie przykładowego pliku HTML w przeglądarce WWW.

Niestety, mimo ogromnego sukcesu pomysłu *World Wide Web*, nie udało się zrealizować wszystkich oczekiwań co do zwiększenia *mocy wyrażania* kodowanych informacji. Wartość pierwotnej koncepcji osłabła paradoksalnie za sprawą znaczników służących osadzaniu grafiki wewnątrz dokumentu. Nową dziedzinę sztuki użytkowej, która się rychło pojawiła, aż nazbyt często można scharakteryzować jako „przerost formy nad treścią”. Słynne zdanie, że „jeden obraz jest więcej wart niż tysiąc słów” brzmi w tym kontekście doprawdy ironicznie i należało by je raczej sparafrazować do postaci „jedno sensownie napisane zdanie warte jest więcej niż megabajty otaczającej je grafiki”. Testowanie kombinacji różnych ustawień znaczników grafiki, wespół z definiowaniem bardzo specjalizowanych tabel w których się je osadza, stało się szybko zasadniczym zajęciem twórców stron internetowych. Umożliwiło to bardzo szybki rozwój komercyjnej strony internetu, a zatem jego upowszechnienie i *de facto* wykreowało oblicze internetu które znamy obecnie, lecz równocześnie pierwotny zamysł, by wzbogacać dokument czysto tekstowy o prosty metajęzyk opisujący jego strukturę nie powiódł się. Na szczęście sama metoda „wzbogacania” przetrwała i ewoluowała dalej, o czym będą traktować kolejne rozdziały niniejszej pracy.

To co stanowi o sile koncepcji WWW — możliwość prostego publikowania w przystępnej formie dowolnych treści — okazało się z czasem być równocześnie jej największą słabością. Twórcy telewizji postrzegali ją przede wszystkim jako medium edukacyjne, w chwili upowszechnienia ich wynalazku, fakt, że informacja stała się towarem i musi konkurować o pozycję rynkową z innymi informacjami z innymi ich wyobrażenie zostało lecz gdy tylko me bardzo szybko jednak przekonali się jak bardzo błędne były ich założenia, gdy informacja stała się towarem.

Przed jedną rzeczą z pewnością internet „broni” się skutecznie. Chodzi o monopolizację – Rumunia, cenzura chińskiego internetu., pucz w Moskwie.

Informacyjna „pajęczyna” teksty obejmuje często również W pierwotnych założenia WWW nie przewidziadotyczyły także który nie przetrwał próby czasu są oczekiwania co do jakości zawartych w internecie informacji. „Sieć” stosunkowo szybko przestała być narzędziem prezentacji wyników prac różnych ośrodków badawczych. Dziś jest po prostu „pajęczyną informacji”, z których niektóre są poznawczo, czy też paradoksalnie artystycznie wartościowe, inne zaś prezentują poziom nad wyraz niski obejmując media, schlebające najniższym biologicznym potrzebą

## 2.6 World Wide Web Consortium

Rysunek 5: Logo World Wide Web Consortium



W roku 1994, w *Massachusetts Institute of Technology*, Tim Berners-Lee dysponując poparciem m.in. CERN, DARPA oraz Komisji Europejskiej, zainicjował działalność organizacji **World Wide Web Consortium** (W3C) [W3Hd] i do dzisiaj przewodzi tej instytucji. Celem funkcjonowania W3C jest „doprowadzenie *Sieci Światowej Skali* do jej pełnego potencjału dzięki rozwijaniu powszechnych protokołów wspomagających jej ewolucję i zapewniających jak najszersze współdziałanie” [W3A].

Konsorcjum skupia parę setek różnych organizacji, wśród których znajdują się ośrodki naukowe, najwięksi producenci sprzętu i oprogramowania komputerowego, szeroka reprezentacja instytucji i stowarzyszeń. Dzięki temu opracowywane standardy mają „otwarty” charakter — powstają drogą ewolucji i konsensusu wśród konkurujących ze sobą często korporacji. Z drugiej jednak strony dość wyraźne jest pragmatyczno-komercyjne ukierunkowanie powstających projektów — wpływy komercyjnego lobby. Negatywne tego skutki mogą polegać na monopolizacji pewnych obszarów technologii, lub na preferencjach dla rozwiązań specyficznych — rozwiązujących konkretny problem, nie zaś rozwiązań ogólnych — pozwalających na szerokie wykorzystanie technologii i stymulujące dalszy jej rozwój.

## 2.7 Przyszłość

Internet rozwija się raczej drogą ewolucji niż rewolucji. Tygiel ludzkich pomysłów konfrontowany jest z brutalną, choć wirtualną, rzeczywistością. Wbrew temu, co mogłoby się wydawać, największe szanse na przetrwanie mają wcale nie te koncepcje, które są w odpowiednim stopniu finansowane. O ile tworzenie infrastruktury internetu wymaga pewnych nakładów, o tyle same narzędzia pozwalające wspierać i rozbudowywać strukturę przetwarzania informacji są dziś relatywnie tanie. Nie dziwi zatem, że znaczący system operacyjny — zasadnicze oprogramowanie każdego komputera — może napisać student<sup>24</sup>, a wirusa komputerowego paraliżującego funkcjonowanie wielu instytucji — licealista. Gdy tylko jakiś pomysł „internetowy” w odpowiednim stopniu dojrzeje, zamykany jest w ramy standardu, tak by mógł znaleźć masowe wykorzystanie. Stopień wykorzystania najświeższych, bardzo abstrakcyjnych standardów dotyczących *Sieci Semantycznej* zależy od tego, czy zostaną one zrozumiane przez szersze grono osób zaangażowanych w rozwój „sieci”, co może być o tyle trudne, że przywołany tu kontekst zagadnień i dziedzin ludzkiej wiedzy jest niezmiernie szeroki.

## 3 *Semantic Web*

Specyfikacje związane z tym przedsięwzięciem można pobrać ze stron internetowych W3C [W3S]. Znajduje się tam następująca definicja:

**Semantic Web** dostarcza ogólnych zrzębów środowiska, które umożliwia, aby **dane** były współdzielone i wielokrotnie wykorzystywane pomiędzy zastosowaniami, przedsięwzięciami i ponad granicami społeczności. *Sieć Semantyczna* jest wspólną inicjatywą prowadzoną pod przewodnictwem W3C z udziałem dużej liczby specjalistów i partnerów korporacyjnych. Bazuje na *Resource Description Framework* (RDF), który integruje wielorakie aplikacje używając XML jako składni i elementów URI dla nazewnictwa<sup>25</sup>.

Po niej zaś ustęp z artykułu opublikowanego na temat *Semantic Web* w *Scientific American* [BLHL01]:

*Sieć Semantyczna* jest rozszerzeniem obecnej *sieci*, w którym informacja zaopatrzona w *dobrze zdefiniowane* znaczenie, pozwala w lepszy sposób pracować wspólnie ludziom i komputerom<sup>26</sup>.

<sup>24</sup>Mam tutaj na myśli system operacyjny *Linux*, który zrodził się w roku 1991 właśnie jako projekt studenta Linusa Thorvaldsa. Wokół tego „otwartego” projektu zgromadziła się bardzo aktywna społeczność. Początkowo bez poparcia dużych instytucji, z czasem z coraz większym ich udziałem *Linux* stał się jednym z powszechniej wykorzystywanych systemów operacyjnych dla serwerów, coraz częściej jest też instalowany na komputerach osobistych.

<sup>25</sup>Przekład autora.

<sup>26</sup>Mając w tym miejscu do dyspozycji polskie wydanie *Scientific American* zdecydowałem się jednak na własne tłumaczenie. Nawet jeśli jest ono językowo mniej poradne, to dość ściśle ujmuje pewne specyficzne terminy. W szczególności *well-defined* (tutaj *dobrze zdefiniowane*) używa się dla określenia poprawności zarówno syntaktycznej (*well-formed*) jak i semantycznej (*well-defined* wymagające *well-formed*) w ramach pewnego języka.



Ten artykuł ma charakter popularyzatorski i tym różni się zdecydowanie od większości dokumentów zamieszczonych na oficjalnych stronach W3C — suchych specyfikacji różnych standardów. Cel autorów w przypadku dwóch cytowanych tekstów jest inny. Czytelnik *Scientific American* zaznajamiany jest przede wszystkim z antycypowanymi możliwościami referowanych technologii, a także konsekwencjami społecznymi ich wykorzystania. Czytelnik stron internetowych W3C to najczęściej specjalista branży IT poszukujący informacji w jaki sposób wykonać konkretne zadanie zgodnie ze standardem. Tak więc dostępna literatura na temat *Semantic Web* obejmuje z jednej strony tematy z zakresu *science-fiction* niemalże, z drugiej zaś ujęcia czysto przedmiotowe. Zdecydowanie brakuje natomiast opracowań, w których analizowano by omawiane zagadnienie na metapoziomie teorii informacji. Ów metapoziom byłby analogiczny do metamatematyki — logiki matematycznej oraz teorii mnogości: metapoziomu dla matematyki. Chciałbym by moja praca w jakimś stopniu wypełniła tę lukę, a zatem muszę odpowiedzieć na pytanie co dla omawianej koncepcji stanowi metapoziom?

### 3.1 Semantyka bez podmiotu poznającego

Logika, lingwistyka, a semantyka w szczególności, lecz nie tylko — wszystkie dziedziny wiedzy rdzenia *cognitive science* z pewnością nadają się doskonale by przeprowadzić z ich pomocą analizę metateorii *Semantic Web*, jednakże z pewnymi zastrzeżeniami. Zasadniczą dla tego nurtu, pochodzącą od Putnama i Fodora *Computational Theory of Mind* [Hor99] należałoby w tym przypadku potraktować jako kontekst, a nie bazę refleksji. Należy położyć nacisk nie na tezę:

**Mózg jest rodzajem cyfrowego komputera.** Stany mentalne korespondujące ze strukturami *języka myśli* są analogiczne do formalnych struktur programów komputerowych, myślenie zaś polega na syntaktycznym — nie semantycznym — operowaniu tymi symbolami.

lecz na tezę następującą:

Jeśli reprezentacje znakowe dowolnego języka powiązane są *pewną relacją* ze *znaczeniami*, to przynajmniej niektóre z tych *znaczeń* poddają się formalizacji umożliwiającej maszynowe ich przetwarzanie (interpretację, wykorzystanie).

Jeśli zgodzić się na niedefiniowanie owej „pewnej relacji”, ani też „znaczenia”, to w dobie badań nad sztuczną inteligencją teza powyższa wydaje się wręcz banalna. I owszem, kontekst tej dziedziny nasuwa się najprędzej jako skojarzenie towarzyszące wyjaśnieniom istoty *Sieci Semantycznej*. Lecz jeśli tylko prawdą jest, że w termin *znaczenie* nie jest wpisane *implicite* tegoż znaczenia przeżywanie (rozumienie) — czynność o charakterze świadomościowym, to nie mamy tutaj bynajmniej do czynienia z koncepcją *strong AI*, gdzie maszyna spełnia test Turinga, lecz co najwyżej z pewnym sposobem *reprezentacji wiedzy* i jej przetwarzania.

Celem przyświecającym twórcom koncepcji *Semantic Web* jest znalezienie takiego sposobu reprezentacji informacji, który pozwoli maksymalnie wykorzystać jej wartość (zawartość) — umieścić ją w określonym kontekście, co z kolei umożliwi *używanie* danych, które ona koduje w dowolnym innym kontekście, w którym jest to potrzebne. Zwycięzają kryteria pragmatyczne — ostatecznie inne być tu nie mogą — z drugiej jednak strony powstały w ten sposób system relacji winien być adekwatny do rzeczywistości<sup>27</sup>. W przeciwnym wypadku cały zamysł nie miałby sensu.

Pytanie o istotę *Sieci Semantycznej* jest zatem pytaniem o specyficzne *rozumienie* terminów takich jak: znaczenie, sens, intensja, referencja, ekstensja, kontekst leksykalno–syntaktyczny, metajęzyk, reguła semantyczna. Pewna językowa nieporadność sformułowania „rozumieć” w tym kontekście wynika z tego, że wyrażenia tu podane zwykle służą jako określenie dla aspektów eksplikacji innych pojęć. Jak to ujął Quine:

Ważne konteksty, w których ludzie mówią (lub sądzą, że mówią) o znaczeniach, sprowadzają się do dwóch: *posiadanie* znaczenia, czyli sensowność, oraz *tożsamość* znaczeń, czyli synonimiczność. Podawanie znaczenia jakiegoś wyrażenia polega po prostu na wymienieniu jego synonimu, sformułowanego zazwyczaj w języku bardziej zrozumiałym niż termin oryginalny. Jeżeli źle znosimy znaczenia jako takie, możemy mówić wprost o wyrażeniach jako sensownych lub nonsensownych i synonimicznych lub heteronomicznych wzajemnie. Problem dostatecznie jasnej i ścisłej eksplikacji przymiotników „sensowny” i „synonimiczny” — najlepiej jak sądzę, w terminach ludzkich zachowań — jest kwestią równie trudną, jak i doniosłą. Natomiast wartość wyjaśniająca nieredukowalnych bytów pośredniczących, zwanych znaczeniami, jest niewątpliwie złudna. [Qui00b, str. 40]

Zatem już sama próba definicji przytoczonych terminów grozi popadnięciem w koło definicyjne polegające na usiłowaniu objaśnienia czym jest „znaczenie znaczenia” lub „sens sensu”<sup>28</sup>. Pozostaje więc jedynie eksplikacja tych wyrażeń dla kontekstu *Sieci Semantycznej* w kategoriach czysto pragmatycznych, lecz nie w „terminach ludzkich zachowań”, jak chciałby Quine (ludzkie „przeżywanie znaczeń językowych”, cokolwiek ono oznacza — jak przekłada się na płaszczyznę behawioralną), tylko według kryteriów sprawności komputacyjnej — możliwości „przejść” między kontekstami użyć (maszynowe, kontekstowe przetwarzanie informacji).

„Pojmowanie” założeń *Semantic Web* musi zatem pociągać za sobą arbitralną — motywowaną kryteriami efektywności, nie zaś dążeniem do ujęcia prawdy — formalizację, i to jak się okazuje nie tylko semantyki, lecz także ontologii, która „wyłania się” jako konsekwencja przyjęcia określonej semantyki [Qui00b]. Komputery pracujące dla człowieka muszą być wyposażone w *system wiedzy*, który da się skutecznie *używać*.

<sup>27</sup>Cokolwiek „adekwatny” tu oznacza (adekwatny do modelu rzeczywistości, adekwatny do świata możliwego). *Nota bene* system ten wewnętrznie koherentny też być musi.

<sup>28</sup>Analogicznie jak definicja analityczności związana jest z odwołaniem do pojęcia definicji, lub synonimiczności. Z kolei synonimiczność da się zdefiniować, jeżeli posiadamy już wcześniej definicję analityczności [Qui00a, str. 62].

Wittgenstein w *Tractatus logico-philosophicus* przedstawił ontologię bez metafizyki [Wit00b], dziś zaś dochodzimy do momentu rozwoju cywilizacji który wymaga od nas tworzenia systemów ontologii, dla których epistemologia nie jest legitymacją. Ontologii dyktowanych ostatecznie kryteriami produktywności — efektywności opisu świata, nie zaś kryteriami prawdziwości tego opisu. Powstaje pytanie, do jakiego stopnia przyjęcie towarzyszącej *Semantic Web* formalizacji semantyki kierowane jest programem „logicznego atomizmu”, do jakiego zaś pragmatycznym ostatecznie rozumieniem *znaczenia*? A może problem ten jest źle postawiony?

## 3.2 Architektura *Semantic Web*

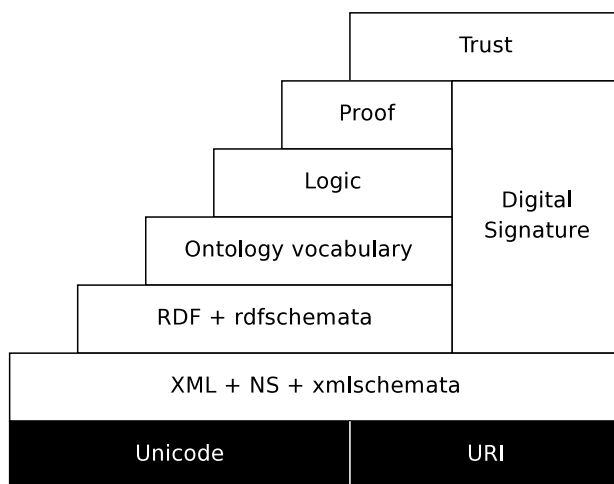
Architektura *Semantic Web* jest czymś w rodzaju wizji — rozważań nad interakcjami które mogą zachodzić pomiędzy poszczególnymi komponentami, czy też warstwami *sieci*. Środki techniczne — języki formalne służące realizacji tej wizji są opisane w rozdziale 3.3 na stronie 33. Niestety o ile środki te w większości już istnieją (jeśli można orzec istnienie w przypadku ich wirtualnego charakteru), to architektura jest w pewnym sensie antycypowana — nie jest ściśle wyznaczonym projektem. W tym przypadku tworzywo wyprzedza w jakimś sensie wizję konstruktora, która wykształca się wraz z budulcem i w pewnym sensie jest nim inspirowana. Analogicznie jak nowe materiały budowlane mogą inspirować architektów. Ewolucyjny charakter rozwoju technologii daje o sobie znać także tutaj.

### 3.2.1 Wieża *Sieci Semantycznej*

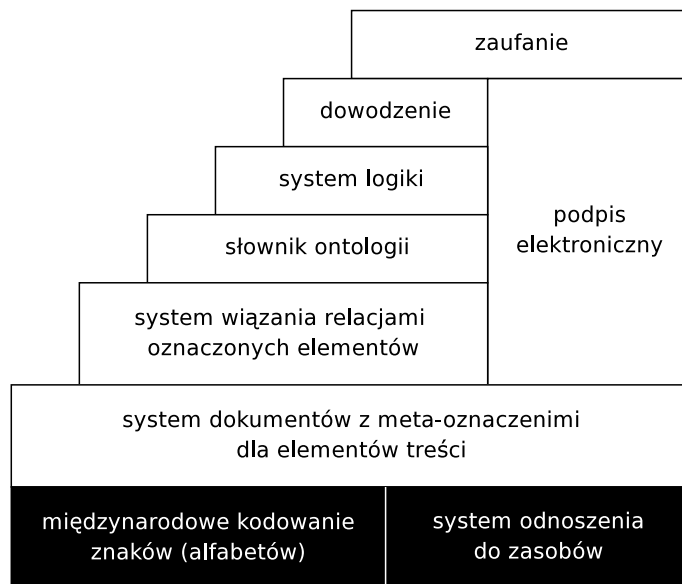
W trakcie powstawania niniejszej pracy miałem duże wątpliwości w jakiej kolejności zaprezentować zgromadzone materiały. Na początku rozdział 3.3 na stronie 33 dotyczący formalnych języków używanych w ramach *Sieci Semantycznej* poprzedzał „architekturę”. Odpowiadało to chronologii, odzwierciedlało proces stopniowego eksplikowania coraz bardziej abstrakcyjnych idei, przy użyciu zdefiniowanych wcześniej środków formalnych. Z drugiej jednak strony postępując ściśle według chronologii musiałbym rozwijać oba rozdziały równolegle. Zwyciężyło ostatecznie przekonanie, by zaczynać od zagadnień ogólnych, a kończyć na szczegółach. Potem odkryłem dokument *The Semantic Web Made Easy* [W3M] który oparto na metodzie pośredniej. Na rysunku 6 znajduje się zaczerpnięta z tego tekstu „wieża *Semantic Web*”. Niestety w poszczególne jej „piętra” wpisane są terminy specyficzne. Przetłumaczona wersja „wieży” uwidoczniła jest na rysunku 7 na następnej stronie.

Na samym dnie umieszczone zostało to, co tylko nieznacznie wykracza poza *warstwę transportową*, a samo w sobie stanowi swoistą „warstwę transportową” dla informacji. Mowa tu o systemie kodowania znaków (*Unicode*), który pozwala na cyfrową reprezentację symboli pochodzących z bardzo dużej ilości światowych języków, także tych ideograficznych (np. *Kanji*).

Równolegle umiejscowiony jest system odnoszenia do zasobów (URI) definiujący reguły nazew-



Rysunek 6: Wieża *Semantic Web*  
(źródło: [W3M])



Rysunek 7: Wieża *Sieci Semantycznej*

nictwa — adresowania zbiorów danych. Umożliwia on jednoznaczne określenie wirtualnego „położenia” dowolnej informacji w ramach *sieci*.

Warstwa ponad nimi to język opisu dokumentów — XML, który pozwala na wzbogacanie ich dowolnie definiowanymi w zależności od potrzeb meta–oznaczeniami. Nowe wymagania mogą owocować powstawaniem nowych szablonów takich oznaczeń, które również w ramach tego „piętra” trzeba uwzględnić (*XML Schema*), a każdy szablon to *de facto* definicja nowego, specyficznego języka. Warto nadmienić, że same szablony również są sporządzane z wykorzystaniem XML.

Przez cztery kolejne poziomy ciągnie się tzw. „podpis elektroniczny”, metoda na uwierzytelnienie treści — dowiedzenie jej prawdziwości, a przynajmniej pochodzenia z kompetentnego źródła.

Do wiązania dowolnymi relacjami elementów specyficznych dokumentów XML (meta–oznaczenia) stosuje się ustanowiony do tego celu „dialekt” XML o nazwie RDF.

OWL, inny „dialekt” XML służy publikacji zbiorów terminów zwanych *ontologiami*. To zazwyczaj tutaj znajdować się będą elementy drugiej strony relacji definiowanej w RDF.

Ponad tym wszystkim znajduje się warstwa mająca na celu ujęcie niższych bazujących na XML elementów i relacji pomiędzy nimi w spójny system logiki formalnej, który umożliwić by miał z kolei utworzenie warstwy wyższej — oprogramowania typu *inteligentny agent*, które potrafiło by wszystko poniżej „potraktować” jako *wiedzę* i przeprowadzić z jej pomocą wnioskowania.

*Zaufanie* na szczycie wieży to określenie na punkt styku między człowiekiem a maszyną która dla niego pracuje. To człowiek musi ostatecznie określić którym źródłem informacji ufa.

Pytanie, czy aby „wieża Sieci Semantycznej” nie zwiastuje nowej „wieży Babel”, pozostawiam w tym miejscu otwarte.

### 3.2.2 Środowisko

Fizycznym środowiskiem *Sieci Semantycznej* jest internet, rozumiany według pierwszej przytoczonej przeze mnie w rozdziale 2.1 na stronie 11 technologiczno–fizycznej definicji. Należy dodać jedynie, że jako węzły do internetu mogą być podłączone nie tylko typowe komputery, lecz dowolnego rodzaju urządzenia elektroniczne. Dokonało się to już częściowo w przypadku telefonów komórkowych i samochodów. Wkrótce czeka to prawdopodobnie także inne urządzenia elektroniczne codziennego użytku, takie jak zegarki, sprzęt audiowizualny, lodówki, pralki, urządzenia kuchenne, etc. W ich przypadku semantyczne oznaczenie dotyczyłoby przede wszystkim cech i funkcji danego urządzenia. Np. telewizor, radio i telefon mogłyby udostępniać informacje o swym rozmiarze wyrażone w centymetrach lub milimetrach, a także o maksymalnym i aktualnym poziomie głośności wyrażonym w decybelach. Innymi słowy chodzi tutaj przede wszystkim o formalne wyeksplikowanie znaczenia dla istniejących funkcji danego sprzętu<sup>29</sup>.

<sup>29</sup>Ciekawe i do tego zabawne przykłady wykorzystania tego rodzaju oznaczeń znajdują się w [BLHL01].

### 3.2.3 Zasoby

Zasoby *Sieci Semantycznej* stanowią dokumenty XML oraz relacje między ich elementami także wyrażone za pomocą XML<sup>30</sup>. W pewnym sensie zasobem są także *ontologie* opisane w rozdziale 3.2.4.

Dokumenty XML przechowują dane, które pozornie nie różnią się pod względem zawartości informacyjnej od danych publikowanych dzisiaj z pomocą HTML<sup>31</sup>. Brak różnicy dotyczy jednak jedynie ludzkiej recepcji treści. Maszyny mogą XML przetwarzać według wielu kryteriów, także tych semantycznych, co jest częstokroć niemożliwe w przypadku HTML.

Warto nadmienić, że dane zawarte w takich dokumentach nie muszą być treściami sporządzonymi przez człowieka, lecz mogą także stanowić sposób prezentacji danych innych systemów. Przykładem może być dynamicznie wygenerowany plik XML opisujący godziny przyjęć lekarzy hipotetycznego szpitala wraz ze specyfikacją ich specjalizacji, stażu, płci, etc.

### 3.2.4 Ontologie

W badaniach nad sztuczną inteligencją termin *ontologia* zyskał inne niż filozoficzne znaczenie. Nie chodzi już o ontologię jako metateorię tego co istnieje, lecz raczej ontologię pojętą po Quine'owsku, jako konkretną, jedną z wielu możliwych odpowiedzi na pytanie o to, co dana teoria „uznaje za istniejące” [Qui00b, str. 44]. Taka *ontologia* stanowi implementację pewnego ogólnego schematu definiowania relacji formalnych pomiędzy terminami — sposób reprezentacji wiedzy. Schematy specyfikowane są tu z pomocą dialektów XML (RDF, OWL)<sup>32</sup>.

**Taksonomia** w kontekście tak pojętej *ontologii* to klasyfikacja obiektów i ich wzajemnych relacji. Jeśli dwa terminy znaczą dokładnie to samo, *ontologia* może zdefiniować je jako ekwiwalentne co pozwala wymieniać je w różnych kontekstach danych. Podobnie jeden termin może być zdefiniowany jako podkategorija (podklasa) innego. Klasa obiektów *Adres* może być zdefiniowana jako podklasa dla *lokalizacja*, zaś klasa obiektów *kod miasta* może być określona w ten sposób, by dało się je przypisać jedynie obiektom należącym do klasy *lokalizacja*. Analogicznie *bliźniak*, *brat* i *siostra* stanowią podkategorię *rodzeństwo*. To oznacza że wszystko co jest prawdą dla kategorii *rodzeństwo* jest również prawdą dla podkategorii *bliźniak*, *brat*, *siostra*.

**Reguły inferencyjne** specyfikują, co może być dowiedzione przez scalanie specyficznych typów informacji o specyficznych kategoriach obiektów. Dostarczają podstaw do przetwarzania informacji publikowanych w ramach *Sieci Semantycznej* przez wyprowadzanie z nich wiedzy. Przykładowa reguła może mieć postać:

Jeśli *kod miasta* jest związany z *kodem stanu* i w *adresie* występuje *kod miasta*, to z *adresem* powiązany jest *kod stanu*.

<sup>30</sup>Struktura XML jest opisana w rozdziale 3.3.1 na stronie 33.

<sup>31</sup>Przykład publikacji HTML znajduje się w rozdziale 2.5.5 na stronie 20.

<sup>32</sup>Opis tych języków znajduje się w rozdziałach 3.3.3 na stronie 38, oraz 3.3.4 na stronie 40.

Dzięki takiej regule możliwe jest przeprowadzenie hipotetycznego wnioskowania:

Skoro siedzibą *Cornell University* jest miasto *Ithaca* to znajduje się ono w stanie *New York*, a zatem w *USA*.

W przypadku tego wnioskowania dane te mogą posłużyć komputerowi do wydrukowania adresu na kopercie w odpowiednim formacie<sup>33</sup>.

*Ontologia* dla Semantic Web stanowi zatem zakodowaną „określoną konceptualizację”, uchwycenie relacji pomiędzy pojęciami. Takie „uchwycenie” nie powinno być dokonywane przez samych twórców stron internetowych (*content provider*). Tylko „globalne” *ontologie* mogą pozwolić na przeprowadzanie „globalnych” wnioskowań, a w tym przypadku inne procesy dowodzenia nie mają sensu. Pojawia się zatem problem zdefiniowania słowników *ontologii* o powszechnym charakterze.

Przywołane w rozdziale 3.2.2 na stronie 29 urządzenia domowe rozgłaszające w ramach *sieci* dane o swoich dostępnych funkcjach i ich stanie musiałyby równocześnie określić czym owe funkcje *de facto* są. Specyficzną dla danego sprzętu (telefon, telewizor, radio) regulację *poziomu głośności* wyrażoną w *decybelach*, należałoby powiązać z odpowiednią globalną *ontologią* dotyczącą sfery *audio*, zaś informacje o rozmiarach z *ontologią* dotyczącą stosunków przestrzennych. Obie *ontologie* powinny wiązać się przy pomocy reguł inferencyjnych z *ontologią systemu miar i wag*.

Dzisiaj takie *ontologie* stopniowo powstają, najczęściej dla dziedzin wiedzy, które przekładają się na zastosowania komercyjne. Największy chyba dzisiaj system tego typu to projekt Cyc firmy Cycorp [CYC], który zawiera w przybliżeniu 3000 terminów pokrywających najbardziej ogólne idee ludzkiej „zdroworozsądkowej rzeczywistości”. Nazwa „Cyc” została „wycięta” ze środka terminu *encyclopedia*.

### 3.2.5 Agenci

John Smith jest Amerykaninem, który postanowił w czasie urlopu spędzić czas wraz z rodziną w górskim kurorcie na „starym kontynencie”. W celu znalezienia konkretnej lokalizacji może skorzystać z dwóch metod. Zgłosić się do biura podróży, lub przeszukać internet.

Niestety, by znaleźć skutecznie coś w dzisiejszym internecie, najczęściej trzeba wiedzieć dokładnie, czym jest to czego się szuka. Wpisanie terminów: „wakacje, Europa, góry” zaowocuje wyświetleniem stron biur podróży o najlepszym „marketingu internetowym”. A zatem powrót do pierwszej możliwości. Wyszukiwanie przynieść by mogło pożądane skutki, gdyby John upatrzył sobie wcześniej jakieś miejsce. Załóżmy że wybór padł na polskie Zakopane. Jeśli John spędzi wystarczająco dużo czasu pracując z wyszukiwarką, dowie się że góry to Tatry, że warto zobaczyć Morskie Oko i zapewne paru innych ciekawostek. Teraz wpisując: „narciarstwo, tatry” i ograniczając wyświetlanie wyników jedynie do stron angielskich znajdzie zapewne w końcu informacje których faktycznie szukał. Niestety ograniczają się one do bardzo konkretnego obszaru terytorialnego.

---

<sup>33</sup>Przykład ten pochodzi z [BLHL01].

Biuro podróży to rodzaj *brokera*. Instytucja taka dysponuje jakąś formą bezpośredniej konceptualizacji pewnego obszaru rzeczywistości, w tym przypadku związanego z ogólnie pojętą turystyką. Stanowi rodzaj „interfejsu komunikacyjnego” pomiędzy nami a ukrytym spektrum możliwości. *Semantic Web* ma być czymś analogicznym, lecz w globalnej skali. Myślę, że w tym aspekcie doskonale charakteryzuje całą koncepcję termin *mega-broker*, choć może i *giga-broker* byłoby określeniem adekwatniejszym.

*Inteligentny agent*, to osobisty program użytkownika internetu, którego celem jest zastosowanie całej warstwowej struktury *Sieci Semantycznej* do przeprowadzenia wyszukiwań i wnioskowań. Jeśli zakodowana w ramach internetu struktura *Semantic Web* ma zostać kiedykolwiek sensownie wykorzystana, to właśnie za sprawą tego typu oprogramowania.

Powróćmy do problemu Johna, lokując go w przyszłości (miejmy nadzieję, że w niezbyt odległej), kiedy dostępne już będą programy — *agenci*. Sposób komunikacji człowieka z jego „pomocnikiem” to temat który od lat eksploatuje literatura *science-fiction*<sup>34</sup>. Dziś trudno powiedzieć na czym będzie on dokładnie polegał. Najprawdopodobniej wykorzystywane zostanie kontekstowe przetwarzaniem języka naturalnego, być może też pojawią się narzędzia pozwalające w sposób graficzny reprezentować abstrakcyjne pojęcia i w jakimś, zapewne mocno ograniczonym zakresie operować nimi.

Spróbujmy wyrazić to czego John potrzebuje:

Znajdź miejscowości w Europie, w których można uprawiać narciarstwo wtedy, kiedy mam wakacje.

*Agent* powinien „wiedzieć” jak owe wakacje przekładają się na *relacje czasowe*. Fakt poszukiwania *lokalizacji* powinien spowodować wykorzystanie reguł inferencyjnych pozwalających przejść od *narciarstwa*, jednej z *form rekreacji*, do *obszarów terytorialnych* w którym taką formę można uprawiać, z uwzględnieniem ich „narciarskości” w odniesieniu do ustalonych na początku *relacji czasowych*.

Wyników takiego wyszukiwania — *miejscowości* leżących wewnątrz wyznaczonych *obszarów terytorialnych*, może być bardzo wiele. W związku z tym John prosi o zawężenie wyników do znajdujących się w tych *miejscowościach hoteli* posiadających *basen*, oraz *pokój* czteroosobowy, gdzie *cena za osobo/dobę* nie przekracza określonej sumy. *Agent* proponuje w tym momencie wymianę kategorii *hotel* na klasę *miejsce noclegowe*, ponieważ *hotel* jest jej podklasą, a pokaźną ilość reprezentantów drugiej podklasy *pensjonat* znaleziono w większości odnalezionych wcześniej miejscowości. Na koniec *agent* może sporządzić kalkulację kosztów takiej podróży. Warto dodać, że w swych poszukiwaniach będzie on często komunikować się z innymi podobnymi sobie programami.

<sup>34</sup>Inspiracje literackie we współczesnej technologii są bardzo żywe. Termin *cyberprzestrzeń* użyty w książce *Neuromancer* Williama Gibsona jest dziś powszechnie stosowany dla określenia conceptualnego, czy też ideograficznego przedstawienia zasobów sieci komputerowej światowej skali. Jest to przestrzeń bytów wirtualnych i ich przedstawień [Gib99].



### 3.3 Środki formalne

Próba kodyfikacji znaczeń, której przyświecają pragmatyczne założenia, prowadzi do powstania nowych, ścisłych języków z określoną specyfikacją. Nie mam zamiaru przedstawiać w tym miejscu żadnych aksjomatów i definicji, lecz pokazać przykłady zastosowań specyficznego metajęzyka do leksykalnej warstwy języka przedmiotowego. Mam nadzieję, że uczyni to opisywane do tej pory zagadnienia nieco mniej abstrakcyjnymi.

#### 3.3.1 *Extensible Markup Language*

Wirtualnym nośnikiem informacji w ramach Semantic Web są **dokumenty**<sup>35</sup>. Stanowią je **pliki tekstowe** wzbogacone o tzw. **znaczniki**. Funkcja znacznika polega na objęciu fragmentu tekstu analogicznie jak robi to cudzysłów w tekście drukowanym<sup>36</sup>. O ile jednak w druku rzadko stosujemy więcej niż jeden rodzaj cudzysłowu, to w pliku wzbogaconym znacznikami możemy wprowadzić dowolne ich typy w dowolnej ilości. Rola znacznika również jest podobna do roli, którą odgrywa tradycyjnie cudzysłów. Stosując go określamy że dany fragment tekstu powinien zostać zinterpretowany w odmienny sposób. Reguły tej interpretacji nie są w żaden sposób związane z samym dokumentem, analogicznie jak kwestią konwencji interpretacyjnej jest rozpoznawanie cudzysłowu<sup>37</sup>.

Opisany format plików — dokumentów, który pozwala na stosowanie dowolnej ilości nazwanych kontekstów cudzysłowowych nosi nazwę *Extensible Markup Language* (XML)<sup>38</sup>. W poniższym przykładzie nazwy i struktura znaczników zostały dobrane w ten sposób, by korespondowały z hierarchiczną strukturą takiego dokumentu jak *Traktat Wittgensteina*<sup>39</sup>.

```
<?xml version="1.0"?>
<tezy>
  <teza numer="1">
    <tekst>
      Świat jest wszystkim co się zdarza.
    </tekst>
  <teza numer="1.1">
    <tekst>
```

<sup>35</sup>Wirtualnym w odróżnieniu od nośnika fizycznego takiego jak dyskietki, twarde dyski i wszelkiego rodzaju „pamięci masowe”.

<sup>36</sup>Przykładowe znaczniki stosowane w HTML zostały opisane w rozdziale 2.5.5 na stronie 20.

<sup>37</sup>Przykładem takiej konwencji mogą być reguły interpretacji cudzysłowu w tekstach polskich i angielskich. W pierwszym wypadku będziemy oznaczali cytaty, w drugim jedną z wypowiedzi dialogu. Jako wyróżnione cudzysłowem rozpoznamy fragmenty tekstu oznaczone jako „tekst” i ”tekst”, a także ’tekst’ oraz «tekst».

<sup>38</sup>Specyfikacja XML dostępna jest na stronach W3C [W3Xa].

<sup>39</sup>W tym przypadku wykorzystałem bardzo trafne, jak się wydaje, modyfikacje, które do przekładu autorstwa Bogusława Wolniewicza[Wit00b] wprowadził przez Wojciech Sady [Wit].

Schematy XML określające sposób kodowania tekstów literackich, a także np. dialogów w języku potocznym opracowywane są przez *Text Encoding Initiative* [TEI]. W tym miejscu zastosowałem własny schemat meta-oznaczeń.

```
    Świat jest ogółem faktów, nie rzeczy.
</tekst>
<teza numer="1.11">
  <tekst>
    Świat jest wyznaczony przez fakty oraz przez to,
    że są to <emfaza>wszystkie</emfaza> fakty.
  </tekst>
</teza>
<teza numer="1.12">
  <tekst>
    Ogół faktów wyznacza bowiem, co się zdarza,
    a także wszystko, co się nie zdarza.
  </tekst>
</teza>
<teza numer="1.13">
  <tekst>
    Światem są fakty w przestrzeni logicznej.
  </tekst>
</teza>
</teza>
<teza numer="1.2">
  <tekst>
    Świat rozpada się na fakty.
  </tekst>
<teza numer="1.21">
  <tekst>
    Jedno może się zdarzyć lub nie zdarzyć,
    a wszystko inne pozostać takie samo.
  </tekst>
</teza>
</teza>
</tezy>
```

Specyficzne formatowanie tego fragmentu tekstu nie jest wymogiem standardu XML. Zostało wprowadzone by ukazać możliwości hierarchicznego, czy w tym przypadku rekursywnego wręcz sposobu stosowania znaczników — elementów do oznaczania treści. By zachować poprawność składniową (*well-formed*) dokumentu wystarczy sygnalizowanie początku i końca znacznika (np. `<teza>` i `</teza>`). Termin *atrybut* został zarezerwowany dla określenia pewnej cechy znacznika, w powyższym przykładzie atrybutem znacznika `<teza>` jest `numer`, a w zależności od pozycji tezy przypisana

jest mu inna wartość.

*XML Schema* [W3Xb] określa w jaki sposób definiować systemy znaczników dla specyficznych zastosowań. W powyższym przypadku konkretny *schemat XML* określałby, że znacznik *teza* może zawierać jeden znacznik *tekst* i żadnego lub nieskończenie wiele innych znaczników *teza*, a także musi mieć przypisany atrybut *numer*. Nowy schemat definiuje *de facto* nowy **format pliku**, a określony dokument w tym formacie jest równocześnie poprawny syntaktycznie w ramach XML (*well-formed*) i *dobrze-zdefiniowany* (*well-defined*).

Jaki jest jednak cel przechowywania tekstu w takiej postaci? Typografia rozumiana jako układ graficzny strony pisma, to zestaw pewnych reguł geometrycznych który koresponduje z określoną konwencją interpretacyjną<sup>40</sup>. Dla czytelnika konwencje są najczęściej „przezroczyste”, a „margines interpretacji” struktury pozostaje zazwyczaj dość wąski. Marginesy, odstępy, wcięcia, interlinie, które w istotny sposób wpływają na nasze rozumienie treści (rozpoznajemy rozdział, akapit, etc.) są zawarte *implicite* w graficznej strukturze stron tekstu. Pisząc *Tractatus* Wittgenstein nie był w stanie wyrazić tego, co miał na myśli, poprzestając na tradycyjnych środkach typograficznych. W sposób arbitralny wprowadził wyrażone *explicite* elementy opisu struktury dzięki zabiegowi specyficznego numerowania tez. W powyższym przykładzie hierarchiczne zależności między tezami są wyznaczone *implicite* przez strukturę użytą do ich zakodowania. Atrybut *numer* w znaczniku `<teza>` jest zatem informacyjnie redundantny, o ile autor nie chciał nadać mu innego, „numerologicznego” znaczenia.

Teraz dysponując odpowiednimi narzędziami możemy zaprezentować ten tekst w dowolnej formie, np. wykresów typu „drzewo” o „korzeniu” odpowiadającym tezie głównej i „gałęziach” prowadzących do podtez, lub też zadać komputerowi pytania dotyczące treści, np. następujące:

Podaj wszystkie tezy zawierające słowo *fakt* które są podtezami tez zawierających słowo *świat*.

Czy słowo *świat* pojawia się w jakiegokolwiek podtezie o stopniu zagłębienia większym niż 2?

Oczywiście takie zapytanie działałoby w przypadku języka angielskiego, w języku polskim niestety trzeba posłużyć się dodatkowymi znacznikami by uwzględnić fleksję:

```
<?xml version="1.0"?>
<tezy>
  <teza numer="1">
    <tekst>
      <klucz nazwa="świat">Świat</klucz> jest wszystkim
      co <klucz nazwa="zdarzać się">się zdarza</klucz>.
```

<sup>40</sup>Takie konwencje mają oczywiście charakter intersubiektywny i różnią się w zależności od kultur, chociażby w ten sposób, że tekst odczytywany jest tradycyjnie począwszy od prawej ku lewej części kartki, przeciwnie — począwszy od lewej, lub też od góry do dołu.

```

</tekst>
<teza numer="1.1">
  <tekst>
    <klucz nazwa="świat">Świat</klucz> jest ogółem
    <klucz nazwa="fakt" liczba="plural">faktów</klucz>,
    <klucz nazwa="rzecz" negacja="prawda" liczba="plural">
      nie rzeczy
    </klucz>.
  </tekst>
</teza>
<teza numer="1.11">
  <tekst>
    <klucz nazwa="świat">Świat</klucz> jest wyznaczony przez
    <klucz nazwa="fakt" liczba="plural">fakty</klucz> oraz przez to,
    że są to <emfaza>wszystkie</emfaza>
    <klucz nazwa="fakt" liczba="plural">fakty</klucz>.
  </tekst>
</teza>
<teza numer="1.12">
  <tekst>
    Ogół <klucz nazwa="fakt" liczba="plural">faktów</klucz> wyznacza
    bowiem, co się <klucz nazwa="zdarzać się">się zdarza</klucz>,
    a także wszystko, co
    <klucz nazwa="zdarzać się" negacja="prawda">się nie zdarza</klucz>.
  </tekst>
</teza>
<teza numer="1.13">
  <tekst>
    <klucz nazwa="świat">Światem</klucz>
    są <klucz nazwa="fakt" liczba="plural">fakty</klucz> w
    <klucz nazwa="przestrzeń logiczna">przestrzeni logicznej</klucz>.
  </tekst>
</teza>
</teza>
<teza numer="1.2">
  <tekst>
    <klucz nazwa="świat">Świat</klucz> rozpada się na
    <klucz nazwa="fakt" liczba="plural">fakty</klucz>.
  </tekst>
</teza>
<teza numer="1.21">
  <tekst>

```

```

    Jedno może <klucz nazwa="zdarzać się">się zdarzyć</klucz>
    lub <klucz negacja="prawda" nazwa="zdarzać się" >nie zdarzyć</klucz>,
    a wszystko inne pozostać takie samo.
  </tekst>
</teza>
</teza>
</teza>
</tezy>

```

W powyższym przykładzie dla *znacznika* <klucz> *atrybuty* liczba i negacja przyjmują w większości wypadków wartości domyślne, odpowiednio singular i fałsz.

Znaczniki <klucz> mogą służyć teraz do automatycznego sporządzenia indexu pojęć. Im bogatszy jest metajęzyk, tym więcej można znaleźć ewentualnych zastosowań, w których dokument XML jest źródłem danych.

Tworzenie tego typu dokumentów w konkretnych formatach powinno być wspomagane przez specjalizowane edytory. Klasyczny komputerowy edytor tekstów funkcjonuje według zasady WYSIWYG — *What You See Is What You Get*. Należało by ją zastąpić regułą WYSIWYM — *What You See Is What You Mean*, „to co widzisz na ekranie jest tym co masz na myśli”.

XML jest specyficzną odmianą *Standard Generalized Markup Language* (SGML), ogólnego projektu określającego sposoby oznaczania treści dokumentów. SGML istnieje od wielu lat, a HTML był jego konkretną implementacją. XML można traktować jako podzbiór SGML, zachowujący cechę produktywności — bycia językiem „specyfikacji dla specyfikacji”.

Niestety, proponowany zarówno w XML jak i w SGML sposób kodowania danych nie nadaje się do opisanego struktury znaczeniowej wszystkich możliwych dokumentów. Za przykład niech posłuży tutaj problem wzajemnego „nachodzenia” tych samych elementów leksykalnych dla różnych sensów, jaki spotkać możemy np. w poezji. Nawet jeśli jesteśmy w stanie określić, co jest ekstensją dwóch metafor, to druga metafora może wykorzystywać część słów pierwszej.

```

<metafora1 ekstensja="moment">mgnienie
<metafora2 ekstensja="enklawa spokoju w centrum chaosu">oka</metafora1>
cyklonu</metafora2>

```

Powyższy fragment XML jest niepoprawny, analogicznie jak niepoprawna byłaby konstrukcja złożona z cudzysłówów o postaci: „mgnienie «oka» cyklonu».

Założenie Fregego, że funkcja sensu zdania określona jest na funkcjach sensu elementów składowych okazuje się w przypadku tego typu zdań języka naturalnego bardzo problematyczne. Rzeczą ciekawą jest, że propozycja nowych języków znaczników dla złożonych dokumentów — *Markup Languages for Complex Documents* (MLCD) — usuwających niektóre wady XML związane z wymogami co do hierarchicznej struktury treści, powstaje w związku z elektronicznym opracowywaniem archiwów Wittgensteina na uniwersytecie w Bergen (Norwegia).

### 3.3.2 *Uniform Resource Identifier*

Wraz z rozwojem internetu pojawiła się potrzeba określenia jednoznacznego sposobu adresowania jego zasobów. *Uniform Resource Identifier* (URI)<sup>41</sup> pozwala wskazywać na wiele rodzajów obiektów, nawet gdy mechanizm dostępu do nich różni się. Oto przykłady:

- `http://www.w3.org/Addressing/`
- `mailto:tomlinson@bbn.com`
- `news:sci.philosophy`
- `http://purl.org/dc/terms/modified`

W pierwszym przypadku odnosimy się do dokumentu W3C opisującego specyfikację URI, w drugim do adresu email wynalazcy poczty elektronicznej, w trzecim do grupy dyskusyjnej na temat filozofii, w czwartym zaś do pojęcia — momentu w którym zasób został zmodyfikowany. Odpowiednia aplikacja — np. przeglądarka WWW, na podstawie URI potrafi „odnieść się” do określonego zasobu. Rozpoznany zostanie protokół komunikacyjny, którego trzeba użyć do pobrania pliku (HTTP), zinterpretowana nazwa serwera na którym taki plik się znajduje (`www.w3.org`) i pobrany plik wskazany przez określoną ścieżkę (`/Addressing/`) — dokument HTML dla strony WWW, opisującej czym jest URI.

URI jednakowoż nie jest zbiorem wytycznych dla maszyny, określających w jaki sposób pobrać z internetu wyspecyfikowany plik. URI to nazwa, identyfikator, zakodowana referencja — to, za pomocą czego odnosimy się do zasobu, obiektu, pewnej ekstensji. Nie zawsze wiadomo do czego URI się odnosi. Mogą to być zarówno zasoby sieciowe, jak i abstrakcyjne obiekty które w sieci nie występują, np. książki, czy też określenie metra jako jednostki miary.

Co więcej, dwa różne URI mogą odnosić do tego samego zasobu i nie ma sposobu na sprawdzenie, że jest to w rzeczywistości ta sama ekstensja. Problem ten może również interferować z opisaną wyżej „niejednoznacznością odniesienia”. Ustalenie, że „Gwiazda Wieczorna” i „Gwiazda Poranna” mają tę samą denotację zajęło ludzkości bardzo wiele czasu i niestety okazuje się, że „globalna informatyka” — komputerowe wnioskowanie, może w przyszłości cierpieć na tę samą agnostyczną przypadłość<sup>42</sup>.

### 3.3.3 *Resource Description Framework*

Choć koncepcja *Semantic Web* wymaga, by jej zasoby posiadały meta–oznaczenia, dokumenty XML stanowią na razie stosunkowo niewielki odsetek plików w internecie. O ile w przypadku XML możemy wprowadzić do dokumentu dowolną strukturę znaczników, to często nikt poza nami, a już z

<sup>41</sup>Specyfikacja dostępna w [W3U].

<sup>42</sup>Zagadnienie to jest opisane w dokumencie *How We Identify Things (on the Semantic Web)?* [Haw].

Rysunek 8: Logo *Resource Description Framework*

całą pewnością komputer, nie wie jakie znaczenia im przypisaliśmy. Na przykład: bardzo często powtarzająca się w nich informacja to dane o autorce. Niestety, w zależności od konkretnych formatów plików XML (*schema*) sposób zakodowania tych danych będzie się różnić.

- `<autor>Janina Kowalska</autor>`
- `<author>John Smith</author>`
- `<entry creator="John Smith">The Matrix has you</entry>`

Zarówno zawartość znaczników `<autor>` i `<author>` jak i atrybutu `creator` w znaczniku `<entry>` należy do tej samej kategorii znaczeniowej. Możliwość „pokazania”, że chodzi o ten sam typ informacji zapewnić ma *Resource Description Framework* (RDF)<sup>43</sup>.

Zgodnie z RDF znaczenie jest kodowane za pomocą trójek uporządkowanych, gdzie kolejne elementy można by określić jako podmiot, orzeczenie i dopełnienie. Oczywiście sam plik zgodny ze specyfikacją RDF również jest poprawnym plikiem XML. Informacja zakodowana w każdej trójce wyraża, że „określony *byt* pozostaje w  *pewnej relacji* z innym *bytem*”. „Byt” w tym przypadku to *klasa bytów* reprezentowana przez dane URI. Analogicznie „pewna relacja” to również URI.

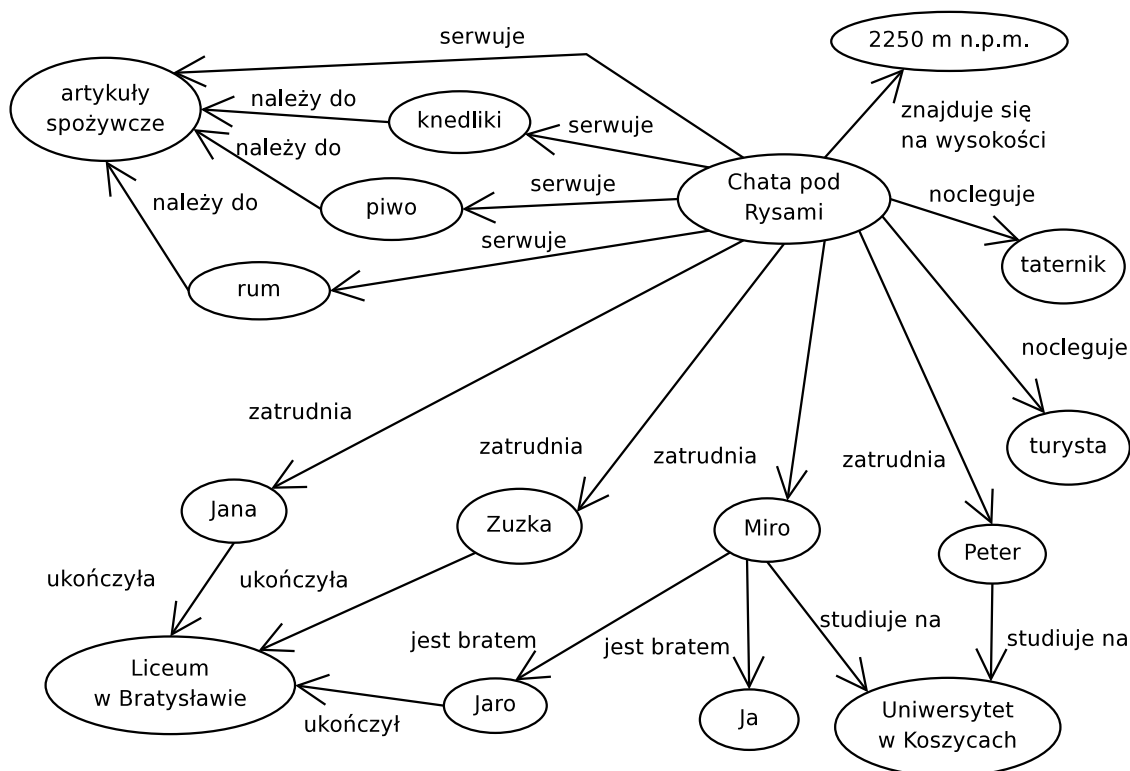
RDF jest zatem sposobem na zakodowanie dowolnych twierdzeń. Możemy określić np., że:

Instancje klasy bytów oznaczonych za pomocą znacznika `<author>` są nazwami dla autora tekstu objętego znacznikiem `<document>`, czyli innymi słowy „author” *jest autorem* „document”.

Określone w ten sposób w RDF relacje można wyrazić za pomocą grafu skierowanego, co przedstawia rysunek 9 na następnej stronie. Jest to graficzna reprezentacja zbioru twierdzeń o świecie — koncepcjonalizacja pewnego jego fragmentu związanego ze schroniskiem górskim w słowackich Tatrach Wysokich. Dzięki RDF tekst dokumentu to nie tylko słowa, lecz także przypisane im pojęcia.

Warto zauważyć także, że model świata wyznaczony przez RDF składa się z „faktów, a nie z rzeczy”. Twierdzenia wiążą abstrakcyjnymi relacjami byty, które posiadają *de facto* naturę wirtualną — są co najwyżej referencjami dla realnych przedmiotów, czy też ich cech. Przy tym taki przedmiot może być wyznaczony całkowicie przez relacje, w jakich znajduje się względem innych elementów. Elementami są tu zarówno przedmioty (obiekty) jak i relacje. Zarówno *podmiotem* jak i *orzeczeniem*

<sup>43</sup>Specyfikacja dostępna w [W3R].



Rysunek 9: Diagram relacji między zasobami

w trójce RDF może być inna relacja. Np. relacja: „ $x$  jest bliźniakiem  $y$ ” może być określona jako *specjalizacja* relacji „ $x$  jest rodzeństwem  $y$ ”. Może zatem formułowane względem Wittgensteina oczekiwanie, by podał przykład zdania atomowego jest niestosowne?

### 3.3.4 Web Ontology Language

10 Lutego 2004 roku W3C zatwierdziło ostatecznie specyfikację RDF i *Web Ontology Language* (OWL)<sup>44</sup>. OWL bazuje na RDF i ma służyć zakodowaniu *ontologii* opisanych w rozdziale 3.2.4 na stronie 30. Zawierają one terminy dotyczące specyficznych dziedzin ludzkiej wiedzy oraz relacje między tymi terminami — twierdzenia na ich temat. Dzięki temu twórczyni zamieszczające w internecie dowolne informacje, będą w stanie za pomocą RDF odnieść je do pewnych globalnych kategorii znaczeniowych, co dopiero umożliwi globalne, maszynowe przetwarzanie danych z uwzględnieniem ich znaczenia.

Specyfikacja OWL wykorzystując RDF rozszerza równocześnie dotychczasową *moc wyjaśniającą* tego standardu o dodatkowe *słownictwo* dla opisanie własności i klas. Dzięki temu można wyrazić np.: relacje między klasami obiektów (np. rozłączność), liczebność (np. „dokładnie jeden”), równoważność, bogatsze typowanie własności, charakterystykę własności (np. symetrię), klasy wyliczeniowe.

<sup>44</sup>Specyfikacja dostępna w [W3O].



Warto tutaj nadmienić, że treści zakodowane w konkretnych formatach XML, a także RDF i OWL, nie muszą być wytwarzane przez człowieka, lecz mogą także być efektem pracy dynamicznych internetowych serwisów, analogicznych do tych opisanych w rozdziale [W3W].

### 3.3.5 *Web Services*

Wiele informacji dostępnych w internecie nie ma charakteru statycznych danych, lecz jest tworzona dynamicznie. Przykładem niech będzie internetowy serwis pogodowy, gdzie udostępniane w nim dane o temperaturze i wilgotności powietrza odczytywane są na żądanie z automatycznych czujników. Wiele tego typu serwisów istnieje, a specyfikacje tak zwanych *Web Services* [W3W] mają za zadanie opisanie jednolitego mechanizmu dostępu do nich, przeznaczonego nie dla ludzi, lecz dla maszyn.

Za przykład *Web Service* może posłużyć kalkulator kursów walut. Specyfikacja danych wejściowych obejmuje kody waluty bazowej, kwotę, oraz kod waluty dla której chcemy obliczyć aktualny kurs. W danych wyjściowych otrzymujemy wartość kursu, oraz zastosowany przelicznik. Gdyby taki *serwis sieciowy* udostępniał także informacje o znaczeniu operacji którą wykonuje, oraz znaczeniu parametrów wejściowych i wyjściowych (RDF, OWL), to automatycznie może on zostać wykorzystany przez oprogramowanie typu *agent* (zobacz 3.2.5 na stronie 31), które byłoby w stanie nie tylko przeprowadzać wnioskowanie, lecz także automatycznie inicjować pewne procesy, a także wykorzystywać wynik ich wykonania.

## 3.4 Przykład zastosowania

Kiedy dzisiaj używamy komputera, by wyszukać w internecie — „wielkiej encyklopedii” — informacje na dany temat, wykorzystujemy procedury podobne do tych, które opisuje Searle w analogii „chińskiego pokoju” [Sea99]. Tradycyjna wyszukiwarka ma za zadanie znaleźć te dokumenty które zawierają wprowadzone w zapytaniu użytkownika słowa. Analogicznie człowiek z „chińskiego pokoju” nie rozumie ani odrobinę ideogramów, którymi operuje<sup>45</sup>. *Semantic Web* ma w zamierzeniu rozwiązać problem owego „rozumienia”.

Założmy, że w przyszłości serwis internetowy Uniwersytetu Szczecińskiego udostępnia informacje o wykładanych przedmiotach i przypisanej im w ramach programu Sokrates punktacji ECTS. Informacje tego typu często się zmieniają. Plik XML może być doskonałym ich nośnikiem, zarówno wtedy, gdy tworzy go człowiek, jak i wtedy gdy dynamicznie generuje go maszyna na podstawie jakiejś istniejącej bazy danych. Oto fragment takiego pliku dla przedmiotów wykładanych na pierwszym roku Filozofii.

```
<przedmiot kod="08.1I10.K01" godziny="60" punkty="7">
```

<sup>45</sup>W przypadku wyszukiwarek można mówić o tym, że „rozumiane” są pewne aspekty frazy zapytania. Np. poszczególne wpisane słowa mogą być porównywane ze słownikiem by określić ich formę gramatyczną, a następnie formę podstawową i w wynikach wyszukiwania uwzględnić także wszelkie formy deklinacyjne lub koniugacyjne. Z pewnością natomiast nie można mówić o przeżywaniu sensu wprowadzonego zapytania, które *nota bene* sensu może nie mieć.

```

    Introduction to Philosophy
</przedmiot>
<przedmiot kod="08.1I10.K02" godziny="60" punkty="7">
    Ancient Philosophy
</przedmiot>
<przedmiot kod="08.1I10.K03" godziny="60" punkty="7">
    Ontology
</przedmiot>
<przedmiot kod="08.1I10.K05" godziny="90" punkty="12">
    Mathematical Logic
</przedmiot>

```

Załóżmy, że nauką w Polsce zainteresowany jest student z innego kraju Unii Europejskiej. Nie wie on jaką uczelnię wybrać, natomiast z pewnością interesuje go filozofia.

Rozpoczyna zatem pracę z własnym *agentem*, który wyszukuje stosowne *uczelnie*. Kryterium jest *prowadzenie zajęć z filozofii*. *Agent* dostarcza listy takich ośrodków, a wyniki mogą być *sortowane według jego szczegółowych zainteresowań* (wnioskowana jest *ekwiwalencja zainteresowań z domenami tematyki badawczej danej uczelni*).

Jeśli doszliśmy z naszym hipotetycznym studentem do etapu przeglądania konkretnych kursów, to możemy prześledzić szczegółowo, na czym polega stosowanie RDF i OWL. Jeśli *punktacja ECTS* dla poszczególnych *przedmiotów* zostałaby skodyfikowana przez odpowiednią jednostkę Komisji Europejskiej w formie opisanej wcześniej *ontologii*, to autorzy stron internetowych Uniwersytetu Szczecińskiego mogą za pomocą RDF powiązać swój specyficzny format pliku z globalnymi kategoriami znaczeniowymi. Wystarczy dostarczyć następujących twierdzeń (kod „@” poprzedza *atrybut*):

- <przedmiot> posiada @kod
- @kod posiada <przedmiot>
- <przedmiot> posiada @godziny
- <przedmiot> trwa @godziny
- <przedmiot> posiada @punkty
- <przedmiot> jest równoważny *course* (ontologia ECTS)
- <przedmiot> zawiera *name of course* (ontologia ECTS)
- @kod jest równoważny *ECTS code* (ontologia ECTS)
- @punkty są równoważne *number of ECTS credits* (ontologia ECTS)

- @godziny są równoważne *number of hours* (ontologia ECTS — *ilość godzin* wyrażona w „godzinach akademickich”)

Dzięki takim twierdzeniom możliwe jest komputerowe porównywanie ofert edukacyjnych europejskich uczelni niezależnie od języka narodowego i systemu danych w którym zostały one zapisane.

Nie tylko śmiałe teorie generują współczesną informacyjną cyberprzestrzeń, lecz także wciąż wzrastający potencjał informacyjny ewokuje, by implementując kolejne generacje systemów przekazywania informacji uczynić je bardziej „rozumnymi”.

## 4 Problemy „semantyki stosowanej”

### 4.1 Która semantyka?

W miarę badania informatycznych aplikacji semantyki ujawnia się bardzo specyficzne ujęcie tej dziedziny. Jest ono z jednej strony zakotwiczone głęboko w tradycji idealistyczno–logicznej (Mill, Frege, Russel, Wittgenstein z czasów *Traktatu*), z drugiej zaś widoczny jest wyraźnie wpływ pragmatycznych koncepcji znaczenia. *Sieć* ma pracować dla ludzi, więc ostatecznie użycie — efektywność wnioskowania, nie zaś „prawdziwość” reprezentacji wiedzy, jest kryterium sensu. Weryfikacja wszelkiej „prawdziwości” nie odwołuje się tu ani do korespondencji, ani nawet koherencji, lecz do „behavioralnego” pragmatyzmu, analogicznie jak „prawdziwa” jest *sztuczna inteligencja*, gdy spełnia test Turinga. Z trzeciej jeszcze strony pojawia się intuicyjny, operacyjny aspekt semantyki. Kodyfikując znaczenie nie pyta się tu co *de facto* znaczy dany leksem, lecz co twórca strony internetowej ma na myśli, używając pewnych ciągów znaków w danym kontekście. Semantyka pojęta w ten sposób związana jest z próbą określenia precyzyjnego aparatu pojęciowego do opisu wzajemnych relacji znaczeniowych dla zjawisk językowych. Tu nie usiłuje się już odpowiedzieć na pytanie: „czym jest znaczenie”, lecz „jak skutecznie je wyeksplikować”.

W ten sposób wyraźnie zbliżamy się do osiągnięć współczesnej lingwistyki, a w szczególności dokonań Chomsky’ego — jej sztandarowego przedstawiciela. Teoria gramatyki generatywnej nie jest w istocie teorią, lecz postulatem, ponieważ analogicznie jak dla psychoanalizy Freuda, nie sposób znaleźć dla tej koncepcji kryteriów weryfikacji. Jednakowoż stanowi ona bardzo nośną „metaforę”, a skoro ustanowieniu *Sieci Semantycznej* przyświecają założenia pragmatyczne, a nie „prawdziwościowe”, zatem każda „skuteczna” metafora przekładająca się na konkretny projekt jest pożądana. Szczególnie interesujące są w tym kontekście późne prace Chomsky’ego związane z rozwinięciem gramatyki generatywnej w semantykę generatywną [Lyo98, str. 87-110]. Mamy tu do czynienia z metaforą „języka jako maszyny”. Znaczenie wyrażenia można traktować jako zbiór jego parafraz — przykład na inną, najlepiej bardziej złożoną sekwencję. Na czym jednakże miałyby polegać kompetencja językowa maszyny?

Równie nośna może okazać się biologiczna metafora „języka jako organizmu”. *Sieć Semantyczna* jest żywa, zmienia się w czasie, a to wiąże się z pytaniem kolejnym: Czy *Semantic Web* to koncepcja zachodniocentryczna? Czy ogólne kategorie znaczeniowe implikujące określoną „sieć znaczeniową” mają charakter obiektywny, czy też są wytworem zbiorowej świadomości językowej — konkretnego paradygmatu — zmiennej logosfery określonej kultury? Faktem jest, że paręset, a nawet dziesiątków lat wcześniej intersubiektywna konceptualizacja rzeczywistości była inna. Za przykład niech posłuży choćby homoseksualizm, który jeszcze 20 lat temu kwalifikowany był w medycynie jako patologia, a dziś jest postrzegany w kulturze zachodniej jako jedna z orientacji. Nic nie wskazuje na to, by w przyszłości nie miały zachodzić zmiany w sposobie kategoryzacji świata. Powstaje pytanie, czy mimo to w dziedzinie semantyczno-lingwistycznej istnieje cokolwiek, co po Kantowsku nazwać by można porządkiem *a priori*? Czy obiektywne schematy kategoryzacji (następstwo, enumeracja, inkluzja, ekwiwalencja, etc.) wpisane w „naturę ludzką” — jedną z najbardziej nadużywanych filozoficzno-etyczno-religijnych kategorii. Jeśli *Sieć Semantyczna* ma być odzwierciedleniem ludzkiej wiedzy, to musi ewoluować, analogicznie jak ewoluuje trzeci świat Poppera. Pytanie tylko, czy istnieje jakiś „darwinizm darwinizmu”, a jeśli tak, to czy w ogóle jesteśmy w stanie w jakikolwiek sposób go określić. Stwierdzenie, że wiedza ma charakter konwencjonalny, zawsze niesie ze sobą jakiś rodzaj aporii, ponieważ zakłada *implicite* własną niesłuszność.

Wydaje się zatem, że charakterystycznemu dla *Semantic Web* ujęciu problemu znaczenia nie towarzyszy implementacja, czy też uproszczenie jakiejś konkretnej teorii znaczenia, lecz że mamy tu do czynienia z rozwiązaniem „eklektycznym”. W imię pragmatycznie pojętej efektywności — możliwości, czy też skuteczności wnioskowań, za to nie zawsze ich poprawności — zaczerpnięto rozwiązania teoretyczne co do znaczenia i rozumienia języka z bardzo wielu szkół.

## 4.2 Informacja, a wiedza

Cel przyświecający realizacji *Semantic Web* w ramach internetu określany jest często jako możliwość przejścia od przechowywania informacji do gromadzenia wiedzy. Być może termin „gromadzenie” jest w tym kontekście bardziej kłopotliwy niż pojęcie „wiedzy”. Czy wiedza może istnieć bez podmiotu poznającego? Czy zgromadzenie jej ma sens w przypadku gdy nigdy nikt z niej nie skorzysta? Czy książka istnieje bez swojego czytelnika, a wiedzą są wszystkie dane zawarte w tabelach matematycznych, a nie tylko te które wykorzystujemy w obliczeniach. Być może należałoby zastąpić termin „gromadzenie” jakimś bardziej dynamicznym odpowiednikiem, takim jak np. „przetwarzanie”. Tak pojęta wiedza ujawnia się zawsze w kontekście dynamicznym — jako możliwość dokonywania wysoce abstrakcyjnych, często rekurencyjnych, często heurystycznych operacji na informacji dzięki dobremu jej zakodowaniu.

## 5 Program badawczy

Praca moja z pewnością nie wyczerpuje postawionego w tytule zagadnienia. Niektóre tematy zdołałem jedynie zaznaczyć. Obawiam się, że wielu nie poruszyłem wcale. Zakres dziedzin wiedzy, które należy uwzględnić przy analizie projektu *Semantic Web* jest ogromny, a przedmiotem mojego zainteresowania w tym miejscu są przede wszystkim te, które wiążą się z jak najogólniej pojętą filozofią. W niniejszym rozdziale chciałbym przedstawić program badawczy — tematy które moim zdaniem warto podjąć w ramach refleksji nad *Siecią Semantyczną* na gruncie czy to tradycyjnych dziedzin filozoficznych, czy też szczegółowych zagadnień w ramach tych dziedzin. W większości wypadków będą to idee typowe dla *cognitive science*. Na końcu każdego podrozdziału postaram się podać literaturę przedmiotu.

### 5.1 Kontekst filozoficzny

Filozofowie wydają się stanowić grupę szczególnie uprawnioną do krytyki (oby konstruktywnej) koncepcji *Sieci Semantycznej*. Potrzebni są dziś specjaliści, którzy uchwycą zależności między pojęciami, tak by stworzyć adekwatne, czy może raczej „użyteczne”, formalne *ontologie*, a właśnie obcowanie z myślą filozoficzną kształtuje intelekt w sposób pozwalający na tworzenie takich modelowych konceptualizacji. Co więcej, pojawia się potrzeba określenia sposobu działania komputerowych narzędzi — najlepiej graficznych — do tworzenia i edycji *ontologii*.

Niestety większość współczesnych prac nad „ontologiami stosowanymi” została wykonana przez inżynierów — specjalistów od komputerów, którzy nie posiadając filozoficznego wykształcenia, często odkrywają ponownie problemy z którymi filozofia zmagala się — często z konkretnymi rezultatami — od setek, czy nawet tysięcy lat.

Ciekawa może się też okazać próba opracowania *ontologii* dla opisu typowych elementów dyskursu filozoficznego. W dokumencie internetowym odnoszącym się do jakiegoś konkretnego stanowiska, twórca strony WWW o metaetyce mógłby odnieść się w specyficznym jej fragmencie do *ontologii* mapującej etyczny kognitywizm, czy też internalizm. Nie chodzi o to, by każdy pogląd zamknąć w ramy typologii, lecz by autor zaznaczył swój sposób rozumienia danego stanowiska, co może ułatwić przeszukiwanie internetu. Przykładowe polecenie dla *agenta* mogło by wyglądać następująco: „Znajdź dokumenty w których Wittgensteinowi przypisywane jest stanowisko solipsystyczne”.

Filozofom nie jest potrzebna szczegółowa znajomość środków formalnych używanych do tworzenia *Semantic Web*, lecz raczej model ich możliwości i sposobów aplikacji. To właśnie była jedna z zasadniczych przesłanek do napisania niniejszej pracy.

## 5.2 Semantyka i lingwistyka

### 5.2.1 Konteksty intensjonalne

Język potoczny obfituje w stosowanie kontekstów intensjonalnych, obfituje w nie zatem nasze potoczne rozumowanie. Czy istnieje sposób przekładu treści illokucyjnych i perlokucyjnych na lokucyjne? Jaki „interfejs użytkownika” winna zapewnić *Sieć Semantyczna*, by nie dopuścić nieścisłości. Jakie illokucyjne i perlokucyjne czynności językowe da się maszynowo zinterpretować?

### 5.2.2 Idealistyczno–sematyczne, czy też pragmatystyczne ujęcie zagadnienia znaczenia?

Problem: czy *Sieć Semantyczna* jest związana z logicznym, czy też pragmatycznym rozumieniem znaczenia językowego, poruszyłem w poprzednich rozdziałach tej pracy wielokrotnie. Tu chciałbym go znowu zaznaczyć, ponieważ wymaga on bardziej szczegółowego opracowania w kontekście omawianych zagadnień.

### 5.2.3 Związki sematyczno–ontyczne

Zagadnienie zaznaczone w tytule tego podrozdziału wiąże się z problemem z rozdziału poprzedniego i podobnie jak ono wymaga systematycznego opracowania. Kontekstu interpretacyjnego mogą dostarczyć z pewnością prace Quine’a.

Bardzo ciekawym tematem wydaje się też opracowanie ontologii Whiteheada i Peirce’a do celów reprezentacji wiedzy w systemach eksperckich.

Literatura:

- <http://www.jfsowa.com/ontology/>

### 5.2.4 Problem wieloznaczności

Już samo stosowanie języka XML ma zapewnić rozwiązanie problemu wieloznaczności. Czy nie jest to jednak rozwiązanie pozorne? Jakie zagrożenia dla poprawności wnioskowań może powodować wieloznaczność?

### 5.2.5 Kontekstualizm i holizm semantyczny

Jednym z rysów charakterystycznych koncepcji *Sieci Semantycznej* nie jest, jak by się mogło zrazu wydawać, próba stworzenia systemu sztucznej inteligencji, lecz maksymalna kontekstualizacja informacji. Idealna kontekstualizacja ma charakter globalny i interkontekstualny, więc jest z kolei paralelna z ideą semantycznego holizmu.

### 5.2.6 Ujmowanie znaczeń za pomocą struktur czysto syntaktycznych oraz ich przekształceń

Jeden z częściej dyskutowanych w ramach *cognitive sciences* problemów. Jakie ograniczenia posiadają współczesne rozwinięcia Leibnizjańskiego „rachunku racjonalnego”.

### 5.2.7 Lingwistyka kognitywna

Lingwistyka kognitywna i metody przetwarzania języka naturalnego mogą pomóc w tworzeniu „interfejsu użytkownika” dla *agentów* i wszelkiego oprogramowania na styku użytkownika *Semantic Web* i „silnika wnioskującego”.

## 5.3 „Z punktu widzenia logiki”

*Semantic Web* ma w zamierzeniach deponować dane istotne dla funkcjonowania społeczeństw. Danym byłyby przypisane bardzo określone znaczenia — komponenty, które należy uwzględnić przy operowaniu tymi danymi, mimo że *sieć* sama w sobie znaczeń „nie rozumie”. Analogicznie logika formalna „nie rozumie” zawartości znaczeniowej twierdzeń, którymi w jej ramach się operuje. *Sieć Semantyczna* musi stanowić właśnie jakiś system logiki, lecz odpowiedź na pytanie: „O jaką logikę tutaj chodzi?” nie jest prosta i jednoznaczna. Czy da się stwierdzić izomorfizm *Semantic Web* z jakimkolwiek znanym systemem algebraicznym? W jaki sposób zagwarantować poprawność wnioskowań dla twierdzeń wkomponowanych w strukturę grafu? Dodatkowym problemem jest otwarty charakter zgromadzonych danych oraz ogromna ich ilość. Wnioskowania mogą być realizowane przez algorytmy o bardzo wysokiej złożoności, a zatem konieczne może być stosowanie heurystyk. W jaki sposób zagwarantować wtedy niepojawianie się sprzeczności we wnioskach, o ile sprzeczność nie była dana w przesłankach? A może należało by do jakiegoś stopnia zaakceptować sprzeczności, a jeśli tak, to jaki jest to stopień?

Tim Berners-Lee jest autorem jednego z niewielu na stronach W3C, za to tematycznie bardzo szerokiego, tekstu odnoszącego się do logiki[BLb]. Dokument ten ma charakter szkicu (*draft*) w ramach sekcji tematycznej „zagadnienia projektowe” (*design issues*)[BLa]. Syntetycznie ujęcie sygnalizowanych w tym artykule problemów da się wyrazić za pomocą pytania: Jaka jest ***moc wyrażania informacji***<sup>46</sup> oraz warunki i skuteczność wnioskowań w ramach różnych opisanych współcześnie systemów logik? Jak więc widać pierwotny dla badań semantycznych Fregego i Russela problem, czy da się traktować język naturalny jako język w sensie logicznym, został tu odwrócony do postaci: „Która z logik najadekwatniej nadaje się do stosowania dla arbitralnie wybranej parafrazy języka naturalnego?”. Przedstawione poniżej zagadnienia zaczerpnięte zostały w większości ze wzmiankowanego artykułu.

<sup>46</sup>Tłumaczenie angielskiego określenia *expressive power*.

### 5.3.1 Aksjomaty

Czy da się sformułować aksjomaty dla wnioskowań w ramach *Semantic Web* w języku logiki pierwszego rzędu? Jeśli tak, to wystarczy wybrać z wielu dostępnych możliwości. Jedną z nich jest na przykład zbiór Nicoda: funktor *nand* (*nie-i*) i reguły dowodzenia w postaci *podstawienia* i *modus ponens*.

Inną propozycją jest logika, w której nie istnieje prawo wyłączonego środka (*horn clause logic*), a która wprowadza równocześnie negację intuicjonistyczną.

Literatura:

- Stanley N. Burris, *Logic for Mathematics and Computer Science*, Prentice Hall 1998.
- <http://thoralf2.uwaterloo.ca/htdocs/stext.html>
- <http://citeseer.nj.nec.com/hogan98putting.html>
- <http://citeseer.nj.nec.com/158146.html>
- <http://www.earlham.edu/~peters/courses/logsys/pnc-pem.htm>
- <http://www.earlham.edu/~peters/courses/logsys/lshome.htm>

### 5.3.2 Access Limited Logic

*Access Limited Logic* (*ALL*) to opisana przez Crawforda i Kuipersa koncepcja systemu reprezentacji wiedzy w postaci sieci. Autorzy dowodzą skuteczność wnioskowań w ramach *ALL*, oraz zalety tej metody w stosunku do bazy wiedzy opartej na rachunku predykatów pierwszego rzędu. Koncepcja *ALL* stanowi znakomity kontekst dla wypracowania logiki przydatnej w ramach *Semantic Web*.

Literatura:

- J. M. Crawford and B.J. Kuipers, *ALL: Formalizing Access Limited Reasoning*, Department of Computer Sciences The University of Texas At Austin Austin, Texas 78712 25 May 1990
- J. M. Crawford and B.J. Kuipers, *Negation and proof by Contradiction in Access Limited Logic*, w *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, Cambridge MA, 1991

### 5.3.3 O RDF i OWL

Dopiero zastosowanie *Resource Description Framework* implikuje system logiki operujący na zbiorach twierdzeń. 10 lutego 2004 wersje dokumentów opisujące semantykę RDF i OWL zostały zaakceptowane jako „rekomendacja” W3C. Znajdują się w nich rozwinięcia zagadnień dotyczących logiki w *Semantic Web*, szczególnie w kontekście teorii modeli.



Oto przetłumaczony fragment jednego z dokumentów<sup>47</sup>:

RDF nie narzuca żadnych logicznych ograniczeń na domeny i zakresy własności; w szczególności własność może być odniesiona do siebie samej. Jeśli RDFy wprowadzą klasy, mogą one zawierać same siebie. Takie „zapętlenie przynależności” jawi się jako złamanie aksjomatu regularności, jednego z aksjomatów standardowej (Zermelo–Fraenkel) teorii zbiorów, która wyklucza nieskończenie zagłębiające się łańcuchy przynależności. Jednakże model semantyczny przedstawiony tu rozróżnia własności i klasy rozumiane jako obiekty od ich *ekstensji* – zbiorów par obiekt–wartość, które spełniają własność, lub rzeczy które są „w” tej klasie – za sprawą dopuszczania rozszerzenia własności lub klasy tak by sama zawierała własność lub klasę bez naruszania aksjomatu regularności. W szczególności to użycie mapowania rozszerzenia klasy pozwala klasom zawierać siebie same. Na przykład: jest całkiem poprawne dla (tego rozszerzenia) „uniwersalnej” klasy, by zawierała tę samą klasę jako swojego członka. Konwencja ta jest często przyjmowana na szczytach hierarchii klasyfikacji. (Jeśli rozszerzenie zawierało samo siebie, wtedy aksjomat może być naruszony, ale ten przypadek nigdy nie zachodzi.) Ta technika jest opisana dokładniej w *A Semantics for the Knowledge Interchange Format*, Hayes, P., Menzel, C., Proceedings of 2001 Workshop on the IEEE Standard Upper Ontology, August 2001.

[...]

Użycie bezpośredniego mapowania rozszerzenia umożliwia, by dwie własności miały dokładnie te same wartości, lub by dwie klasy mogły zawierać te same instancje i wciąż być różnymi encjami (bytami). To oznacza, że klasy RDF-owe mogą być pojmowane jako będące czymś więcej niż tylko prostymi zbiorami; mogą być pomyślane jako „klasyfikacje” lub „koncepty” (pojęcia) które mają mocną notację identyczności, wykraczającą poza prostą ekstensjonalną korespondencję. Ta własność teorii modeli ma znaczące konsekwencje w językach zbudowanych za pomocą RDF, a o większej mocy wyrażania, takich jak OWL, gdzie możliwe jest wyrażenie identyczności zarówno pomiędzy własnościami jak i klasami w sposób bezpośredni. Ta „intensjonalna” natura klas i własności jest czasem określana jako użyteczna własność języka deskryptywnego. Pełna dyskusja zagadnienia wykracza poza zakres tego dokumentu.

Warto zauważyć, że pytanie, czy klasa zawiera samą siebie jako członka, jest zupełnie różne od pytania, czy jest ona podklasą siebie samej. Wszystkie klasy są własnymi podklasami.

Literatura:

- <http://www.w3.org/TR/rdf-mt/>

<sup>47</sup>[Hay], przekład autora.

- <http://www.w3.org/TR/owl-semantic/>

#### 5.3.4 Logiki wyższych rzędów

Wiele przemawia za tym, że logika klasyczna nie jest w stanie wyrazić typowo ludzkiego sposobu wnioskowania w realistyczny sposób. Z drugiej strony „nie ma dobrego zbioru aksjomatów i reguł dla logik wyższych rzędów” (*Higher-Order Logic*).

Logiki wyższych rzędów pomagają opisać relację między inną relacją a obiektem, lub relację między dwoma innymi relacjami, więc mogą być bardzo przydatne jako model logiki *Sieci Semantycznej*. Możliwość zapisania kompilatora danego języka programowania w tymże języku jest zazwyczaj pożądaną cechą — niejako testem dojrzałości. Takim testem dla określonego modelu logiki byłoby opisanie go z pomocą jego własnych reguł.

Literatura:

- Jaroslav Peregrin, *What Does One Need, When She Needs Higher-Order Logic?*, Proceedings of LOGICA'96, FILOSOFIA, Praha, 1997, 75-92  
<http://dec59.ruk.cuni.cz/~peregrin/HTMLTxt/hol.htm>

#### 5.3.5 Indukcja, prymitywna rekursja i generalizacja na nieskończenie wiele przypadków

Logika pierwszego rzędu nie jest wystarczająca do przeprowadzenia wnioskowania we wszystkich możliwych przypadkach. W niektórych z nich potrzebny jest jakiś rodzaj indukcji wraz z formalizującą ją teorią, a to najprawdopodobniej pociąga za sobą konieczność odwołania się do logiki wyższego rzędu.

Literatura:

- Robert S. Boyer  
<http://theory.stanford.edu/people/uribe/mail/qed.messages/22.html>
- Peter Suber, *Recursive Function Theory*  
<http://www.earlham.edu/~peters/courses/logsys/recursiv.htm>
- <http://www.w3.org/2001/03swell/pra.n3>
- <http://www.cs.utexas.edu/users/moore/publications/km97a.ps.Z>
- <http://rdfig.xmlhack.com/2002/03/26/2002-03-26.html#1017177958.271019>
- McCarthy Circumscription  
<http://www-formal.stanford.edu/jmc/circumscription.html>

### 5.3.6 Logiki parakonsystentne

Jak wynika z rozdziału 5.3.1 logiki parakonsystentne mogą okazać się przydatne jako model wnioskowania dla *Sieci Semantycznej*. Należałoby zatem zbadać różne opisane logiki tego rodzaju pod kątem ich mocy wyrażania informacji (*expressive power*).

### 5.3.7 Rozstrzygalność i problemy nierozstrzygalne

O ile koncepcja *Semantic Web* zakłada spójność twierdzeń, na bazie których będzie przeprowadzane wnioskowanie, to jednym z jej założeń projektowych jest możliwość wyrażania stwierdzeń nierozstrzygalnych i trudnych (*intractable*) problemów. Tim Berners-Lee podaje za Crawfordem i Kuiperssem<sup>48</sup>:

Doświadczenie z formalnie wyspecyfikowanymi systemami reprezentacji wiedzy dowodzi, że istnieje związek między mocą wyrażania informacji (*expressive power*) za pomocą tych systemów a ich złożonością obliczeniową<sup>49</sup>. Jeśli, dla przykładu, system reprezentacji wiedzy ma moc wyrażania informacji na poziomie rachunku predykatów pierwszego rzędu, wtedy problem określenia co *agent* może logicznie wywnioskować ze swojej bazy wiedzy jest nierozwiązywalny.

Związany z nazwiskiem Göedla problem rozstrzygalności i jej wpływu na wnioskowanie wymaga dokładnego zbadania w kontekście *Sieci Semantycznej*.

### 5.3.8 Ujęcie metajęzyka w ramach języka przedmiotowego?

Interesujące są rozwinięcia wyników badań Tarskiego odnośnie meta-deskrypcji. Jeśli spójny język nie może mieć tej samej mocy wyrażania treści co jego własna metateoria, to jakie mogą być tego konsekwencje dla wykorzystania takich języków jak np. XML? Inny kontekst stanowi tu paradoks Montague'a — nawet proste języki nie potrafią spójnie opisać własnej semantyki.

Literatura:

- <http://lists.w3.org/Archives/Public/www-archive/2002Apr/0057.html>

### 5.3.9 Ujęcie elementów negatywnych

W jaki sposób wyrazić *negatywne stany rzeczy*? W jaki sposób zakodować negację w dokumentach XML? W jaki sposób wyrazić ją w *ontologiach*? Jeśli *funktor negacji* na poziomie języka potocznego ma charakter inny niż w *logice klasycznej*, to w jaki sposób przeprowadzić wnioskowanie? W jaki sposób stosować negację we wnioskowaniu, dowodzić nie wprost?

Literatura:

<sup>48</sup>Zaczerpnięte z [BLb], tłumaczenie autora.

<sup>49</sup>Złożoność obliczeniowa to termin z zakresu teorii algorytmów.

- Z. Kowalski, W. Krysztofiak, A. Biłat: *Koncepcje negatywnych stanów rzeczy*, Wyd. UMCS: Lublin 1998

### 5.3.10 Czy *Semantic Web* to nowa sylogistyka?

Bardzo krytyczne ujęcie zagadnienia *Sieci Semantycznej* przedstawił Clay Shirky.

Literatura:

- [http://www.shirky.com/writings/semantic\\_syllogism.html](http://www.shirky.com/writings/semantic_syllogism.html)

## 5.4 Współczesny „język idealny”?

Oto niepełna lista autorów, którzy opracowywali różne koncepcje języka idealnego — uniwersalnego:

- Raymond Lull, (XIII w.) *Ars Magna*
- Rene Descartes, (1596–1650), *Mathesis Universalis*
- Gottfried Leibniz, (1642–1727), *Rational Calculus (calculus ratiocinator)*
- George Boole, (1815–1864), *The Laws of Thought*, Algebra Boole’a, logika formalna
- G. Frege, A.N. Whitehead, B. Russell
- Kurt Göedel
- Noam Chomsky

Poszczególne próby sformułowania języka uniwersalnego miały różny charakter, lecz z pewnością koncepcja *Semantic Web* wpisuje się w ten ciąg znakomicie, choć nie ma to już być język przeznaczony dla ludzi, lecz przede wszystkim język maszyn, który ma pomóc dopiero ludziom w skutecznej z maszynami interakcji. Opracowanie tego tematu wiązałoby się z badaniami dorobku minionych epok, pod kątem możliwości jego skutecznego wykorzystania dzisiaj.

Literatura:

- Jaap Maat, *Philosophical Languages in the Seventeenth Century: Dalgarno, Wilkins, Leibniz*, University of Amsterdam
- Umberto Eco, *The Search for the Perfect Language (The Making of Europe)*, Blackwell Publishers
- Umberto Eco, *Serendipities: Language and Lunacy*, Harvest Books 1999

## 5.5 Psychologia, antropologia

Badania psychologiczne są cenne dla rozwijania *Sieci Semantycznej* o tyle, że mogą pomóc w rozumieniu sposobów, w jaki kształtują się zdolności poznawcze człowieka, a zatem również zdolność myślenia abstrakcyjnego i kategoryzacji.

Czy kontekst kulturowy może determinować rodzaj wiedzy? Problem ten w wersji lingwistycznej określa hipoteza Sapira–Whorfa. Czy zachodni *logos* jest zmaskulinizowany — czy istnieje feministyczna epistemologia? Jeśli tak, to studia *gender* powinny inspirować specjalistów od sztucznej inteligencji.

Literatura:

- Alison Adam, *Artificial Knowing: Gender and the Thinking Machine*

## 5.6 Filozofia umysłu, neurobiologia i ewolucjonizm

Badanie możliwości poznawczych zwierząt — nie tylko psychologiczne, czy behawioralne, ale też neurologiczne — ujawnia zupełnie nowe perspektywy dla zagadnienia wiedzy, jej gromadzenia i przetwarzania. Wyniki tych badań powinny być kontekstem, w którym *Sieć Semantyczną* należy wciąż weryfikować.

Tytułowy bohater książki *O mężczyźnie który pomylił swoją żonę z kapeluszem*, najprawdopodobniej z dość szczególnym uszkodzeniem mózgu, w sposób poprawny analizował, wnioskował, wręcz algorytmizował, lecz intuicyjne, syntetyczne ujęcie sensu przedmiotu było mu niedostępne w doświadczeniu obrazowym (agnozja). W róży widział możliwą „symetrię wyższego rzędu”, nie zaś kwiat, dopóki jej nie powąchał [Sac96, str. 26-42]. Jeśli nasza wiedza ma charakter obrazowo–dźwiękowo–węchowy, nie zaś językowy czy „obliczeniowy” (komputacyjny), jeśli jest rodzajem sztuki, to jaka jest mentalna reprezentacja języka? Zarówno reprezentacja języka mówionego, który zdaniem lingwistów ma prymat nad pisanym, jak i tego drugiego — obrazowo–graficznego właśnie.

### 5.6.1 Semantic Networks

*Semantic Network* to termin użyty po raz pierwszy przez Rossa Quilliana w roku 1968 w celu opisu sposobu, w jaki umysł ludzki zapamiętuje znaczenia. Rozwijana jest tu koncepcja sieci asocjacyjnej. Teoria ta jest wciąż rozwijana, a *Semantic Web* można potraktować z pewnymi zastrzeżeniami jako jedną z prób jej zamodelowania. Inne implementacje związane są z rozwiązaniami problemu gromadzenia wiedzy w istniejących systemach eksperckich (przykładem może być system SemNet [Sem]).

Literatura:

- M. Ross Quillian, *Semantic Memories*, w: M. M. Minsky, editor, *Semantic Information Processing*. Cambridge, MA: MIT Press, 1968

- *Searching Intranet Using Knowledge Representation – Experiences with the SearchEnhancer*  
<http://w3.informatik.gu.se/~dixi/publ/snet.htm>

### 5.6.2 Sieci neuronowe

Sieci neuronowe z jednej strony mogą być wykorzystane jako metoda wnioskowania (analogon heurystyki) dla *agentów* w sytuacji niepewności, z drugiej zaś strony można szukać analogii między strukturami neuronowymi i bardziej ogólnymi „pojęciami” które z ich pomocą się ujmuje, a siecią powiązań znaków (nośnikiem XML jest ostatecznie plik binarny) i przypisywanych im znaczeń w ramach *Semantic Web*. Może to być metafora dla szczególnego rozumienia *mind–body problem*.

## 5.7 Systemy eksperckie

Metody wnioskowania operujące na *Sieci Semantycznej* jako *bazie wiedzy* będą w większości analogiczne do tych stosowanych w ramach systemów eksperckich, a w szczególności do tych pośród nich, które bazują na *semantic networks*<sup>50</sup>.

W ramach *Semantic Web* nie definiuje się systemu reguł wnioskowania, natomiast określa się jakie operacje na *bazie wiedzy* są dopuszczalne i poprawne w ramach danego kontekstu odniesień. *Sieć* sama w sobie nie dostarcza narzędzi do zadawania pytań, czy też do sprawdzania poprawności twierdzeń. Z punktu widzenia teorii systemów eksperckich jest sposobem reprezentacji wiedzy.

Oprogramowaniem, które obsługuje proces wnioskowania, jest najczęściej *agent*<sup>51</sup>.

Interesujące zagadnienia za zakresu systemów eksperckich to:

- systemy reprezentacji wiedzy
- heurystyki
- uczenie maszynowe

Literatura:

- John F. Sowa, David Dietz, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, 1999

## 5.8 Sztuczna inteligencja

*Semantic Web* nie nawiązuje bezpośrednio do koncepcji sztucznej inteligencji. Specyfikację standardu W3C można traktować jako bazę wiedzy, a to jest termin typowy dla systemów eksperckich, a więc zagadnień kwalifikowanych tradycyjnie do dziedziny tzw. *słabej AI* (ang. *weak AI*). *De facto* wiele

<sup>50</sup>Opisanych w rozdziale 5.6.1 na poprzedniej stronie.

<sup>51</sup>Wymagania funkcjonalne stawiane programom typu *agent* zostały opisane w rozdziale 3.2.5 na stronie 31.

z problemów opisanych w ramach rozdziału „Program badawczy” rozważa się w ramach badań nad sztuczną inteligencją. Tak więc tematyka związana z *Artificial Intelligence* jest po części wspólna z tematyką *Semantic Web*, a po części te dziedziny stanowią dla siebie nawzajem bardzo solidne konteksty interpretacyjne. Wymagania stawiane w przypadku *Sieci Semantycznej* dla AI są mniejsze — raczej nie trzeba rekonstruować kontekstu w ramach którego dokonuje się wnioskowanie — model świata w postaci różnych ontologii jest dany. To tak jakby posiadać absolutną pamięć, a potrzebować jedynie czasu, być może nieskończonego, na znalezienie stosownych danych. Zagadnienia AI z całą pewnością dotyczą w wielkim stopniu *agentów*.

Tak zdecydowane dawniej rozróżnienie w ramach badań nad sztuczną inteligencją na wersję *strong* i *weak*, wraz z rozwojem *cognitive sciences* jest wypierane przez teorię *computational intelligence* — ujęcie unifikujące.

## 5.9 Wykładanie koncepcji

Twórcy *Semantic Web* mają globalne aspiracje i trudno się temu dziwić, ponieważ bez szerokiego zastosowania ta koncepcja nie ma sensu. Pojawia się zatem problem: w jaki sposób ją spopularyzować? Sprawę dodatkowo komplikuje fakt, że potencjalni odbiorcy — „użytkownicy” tej idei dzielą się na następujące grupy:

### 5.9.1 Programiści i specjaliści IT

Twórcy kolejnych standardów przechowywania i przekazywania informacji mają zazwyczaj bardzo praktyczne cele, do osiągnięcia których nie potrzebna jest im zbyt „wyrafinowana” podbudowa teoretyczna. W kategoriach profesjonalnych sytuują się więc, o ile można to tak ująć, gdzieś między rzemieślnikiem a architektem. By zdefiniować nowy, prosty język, niepotrzebna jest bardzo złożona wiedza o lingwistyce. Z drugiej jednak strony wyzwanie uczynienia komunikacji maszyn bardziej „rozumną” prowokuje, by zgłębić poziom metateorii owej praktycznej działalności. Dochodzi zatem do sytuacji w której sfera *praxis* musi z konieczności sięgnąć do *episteme* i to właśnie do tego zakresu *episteme*, który bada warunki ludzkiego poznania — wewnętrzny język podmiotu.

Programiści czy też ogólnie specjaliści branży IT stanowią do tej pory najszersze grono zainteresowane różnymi aspektami *Semantic Web*. Celowo podkreślam fakt, że ich zainteresowania są najczęściej aspektowe. Z pewnością z pożytkiem byłoby stworzenie popularyzatorskich materiałów poruszających tematykę specyficzną dla *cognitive science*, a kierowanych właśnie do tej grupy odbiorców.

### 5.9.2 Twórcy treści

Drugą grupą ludzi, wśród których omawiana tematyka winna być spopularyzowana, są **twórcy treści**<sup>52</sup>. Nie chodzi tutaj o żadną szczególną profesję, lecz o tych, którzy zajmują się dostarczaniem danych do internetowych dokumentów — dziennikarzy, pracowników różnych instytucji, naukowców, a także zwykłych użytkowników „sieci”. Środowiskiem ich pracy powinny stać się programy — edytory tekstu pomagające wprowadzać oznaczenia semantyczne. Do obsługi takiego oprogramowania potrzebna jest stosowna teoria, którą należy opracować.

### 5.9.3 Popularyzatorzy

Do nich między innymi kierowana jest niniejsza praca. Chodzi o ludzi, którzy dzięki swoim psychologicznym i pedagogicznym zainteresowaniom skłonni byłiby przełożyć abstrakcją koncepcję *Sieci Semantycznej* na język czytelnych metafor i analogii.

## 5.10 Oprogramowanie

Zakres typów oprogramowania korzystającego potencjalnie z *Semantic Web* jest ogromny. Wydaje się wręcz, że o ile koncepcja ta zostanie urzeczywistniona chociażby w niewielkim zakresie, to programy nie oferujące interfejsu dostępu do zasobów *sieci* staną się mało atrakcyjne, analogicznie jak trudno mówić o „dobrym” funkcjonowaniu współczesnego komputera osobistego, jeśli nie ma on dostępu do internetu.

## 6 Zakończenie

Analiza standardów, które w przyszłości mają konstituować *Semantic Web*, nie powinna odbywać się w „akademickiej próżni”, ponieważ wyniki tego rodzaju badań mogą okazać się niezmiernie interesujące dla specjalistów wielu dziedzin. Badania nad abstrakcyjnymi strukturami, inspirowane są bardzo konkretnymi potrzebami. Ów pragmatyzm może częstokroć filozofów odstraszać. Termin „technologia”, odniesiony do filozofii, ma w dużej mierze pejoratywne konotacje, a z pewnością właśnie ten termin celnie charakteryzuje rodzaj podejścia reprezentowanego często w ramach nurtu czy to filozofii analitycznej czy też *cognitive science*. Z drugiej strony zainteresowanie ze strony „praktyków” — „technologów” konkretnych dziedzin dla metateorii ich działalności jest częstokroć bliskie zeru. To tak jakby fizyk teoretyk nie interesował się wynikami prac fizyka eksperymentatora, i *vice versa*<sup>53</sup>.

<sup>52</sup>W języku angielskim istnieje określenie *content provider*, co jest zazwyczaj tłumaczone na język polski jako „dostawca zawartości”. Być może właśnie dzięki *Semantic Web* zyska na prestiżu treść — sens komunikatów językowych, ich lokucyjne, nie zaś illokucyjne aspekty. Dlatego wolę nieco wykraczające w przyszłość (niestety) określenie „twórcy treści”.

<sup>53</sup>Poruszyła mnie jakiś czas temu informacja opublikowana w *Świecie Nauki*, polskim wydaniu *Scientific American*, że pewne wyniki fizyków teoretycznych stały się inspiracją dla badań fizyków eksperymentalnych. Badania te niestety wy-



Odstraszać może także konwencjonalny charakter analizowanych struktur, ale czy z drugiej strony nie są one w takim samym stopniu konwencjonalne jak język, w którym jest wyrażona chociażby ta praca? Aż chciałoby się rzec: „w pragmatycznym kryterium sensu tkwi zaiste coś Prawdziwego”. Drogą rozwoju technologii jest ewolucja, nie należy jednak nie doceniać znaczenia mutacji, które często pociągają za sobą jakiś rodzaj małej rewolucji. Ewolucja ewokuje i jest ewokowana — posiada „sprzężenie zwrotne”<sup>54</sup>, ale istnieje też coś w rodzaju nowego bodźca, przekroczenia pewnego „punktu krytycznego” — czasem wręcz zmiany Kuhnowskiego paradygmatu. Jeśli konsekwentnie przykładamy analogie ewolucyjne do kwestii „rozwoju myśli”, to człowiek staje się „komórką”, a jego świadomość paradoksalnie „nośnikiem DNA” dla *memów*, czy też *semów*. W tym sensie to nie ludzie stworzyli internet, lecz *memy* z pomocą mutacji własnego „organizmu”, zbioru wszystkich „komórek”, zaadaptowały sobie nowe środowisko życia. Jeśli jest w tym ujęciu coś słusznego, to są one doprawdy bardzo „samolubne”.

Mając powyższe na uwadze dochodzę do wniosku, że przedstawiony w rozdziale 5 program badawczy musi „powrócić do własnych korzeni”. Chciałbym, by opracowanie tych analiz znalazło się z powrotem w internecie, dostępne każdemu kto znajdzie w nim coś interesującego — zarówno w wersji polskiej jak i angielskiej. Oczywiście sposób prezentacji treści oraz ich zakodowania w ramach takiego internetowego serwisu związany by był z maksymalnym wykorzystaniem obecnych standardów *Sieci Semantycznej*, a równocześnie doskonale nadawałby się do ich testowania i mógłby w przyszłości stać się forum dla wszystkich chętnych zainteresowanych tym tematem którzy chcieliby taki serwis współredagować. Byłoby to ujęcie treści na temat sposobów ujmowania treści, za pomocą tych właśnie sposobów. A oto URL tego przyszłego serwisu:

<http://www.xemantic.com/>

---

kazały niezgodność praktyki z teorią. Fizycy teoretyczni zaapelowali o powtórzenie badań i tym razem wyniki teoretyczne i praktyczne zgadzały się. Poprzednie rozbieżności zrzucano na karb błędnego funkcjonowania aparatury pomiarowej. Chciałoby się powiedzieć za Heglem: „tym gorzej dla faktów”.

<sup>54</sup>Jest to jeden z ulubionych terminów „cybernetycznych” Stanisława Lema, a jego *Summa Technologiae* doskonale pokazuje w jaki sposób można odnieść kryteria ewolucyjne do technologii.

## Literatura

- [Bac99] Kent Bach, *Meaning*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [Bad99] Alan Baddeley, *Memory*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [BLa] Tim Berners-Lee, *Design issues, architectural and philosophical points*, <http://www.w3.org/DesignIssues/>.
- [BLb] ———, *The semantic web as a language of logic*, <http://www.w3.org/DesignIssues/Logic.html>.
- [BLc] ———, *Tim berners-lee*, <http://www.w3.org/People/Berners-Lee/>.
- [BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila, *The semantic web*, 34-43 **5** (2001), no. 284, *Scientific American*, <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>.
- [CYC] *Cycorp, inc.*, <http://www.cyc.com/>.
- [Gal90] Charles R. Gallistel, *The organization of learning*, A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England, 1990.
- [Gib99] William Gibson, *Neuromancer*, przeł. Piotr W. Cholewa, Zysk i S-ka, 1999.
- [Haw] Sandro Hawke, *How we identify things (on the semantic web)?*, <http://www.w3.org/2001/03/identification-problem/>.
- [Hay] Patrick Hayes, *Rdf semantics*, <http://www.w3.org/TR/rdf-mt/>.
- [Hel00] Michał Heller, *Cywilizacja „czarnej skrzynki”*, Tygodnik Powszechny (2000), no. 17.
- [Hor99] Steven Horst, *Computational theory of mind*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [HW99] Laurence Horn and Gregory Ward, *Pragmatics*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [ISC] *Internet systems consortium*, <http://www.isc.org/>.
- [Kry] *Notatki z kursu „semantyka” prowadzonego przez dra wojciecha krysztofiaka w roku akademickim 2002/2003.*

- [LCC<sup>+</sup>] Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, and Stephen Wolff, *A brief history of the internet*, <http://www.isoc.org/internet/history/brief.shtml>.
- [Lyo98] John Lyons, *Chomsky*, przeł. Barbara Stanosz, Prószyński i S-ka, Warszawa, 1998.
- [Par99] Barbara H. Partee, *Semantics*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [Pri] *Global consciousness project*, <http://noosphere.princeton.edu/>.
- [Qui00a] Willard Van Orman Quine, *Dwa dogmaty empiryzmu*, Z punktu widzenia logiki. Dziewięć esejów logiczno-filozoficznych, przeł. Barbara Stanosz, Aletheia, 2000, pp. 49–75.
- [Qui00b] ———, *O tym co istnieje*, Z punktu widzenia logiki. Dziewięć esejów logiczno-filozoficznych, przeł. Barbara Stanosz, Aletheia, 2000, pp. 29–47.
- [Sac96] Oliver Sacks, *Mężczyzna, który pomylił swoją żonę z kapeluszem*, Zysk i S-ka, 1996.
- [Sea99] John R. Searle, *Chinese room argument*, The MIT Encyclopedia of the Cognitive Sciences (Robert A. Wilson and Frank C. Keil, eds.), MIT Press, 1999.
- [Sem] *About the semnet software*, <http://www.biologylessons.sdsu.edu/ta/about/aboutsemnet.html>.
- [Squ99] Larry Squire, *Memory, human neuropsychology*, The MIT Encyclopedia of the Cognitive Sciences (MITECS), MIT Press, 1999.
- [TEI] *Welcome to the tei website*, <http://www.tei-c.org/>.
- [W3A] *About the world wide web consortium (w3c)*, <http://www.w3.org/Consortium/>.
- [W3Ha] *Http - hypertext transfer protocol*, <http://www.w3.org/Protocols/>.
- [W3Hb] *Hypertext markup language (html) home page*, <http://www.w3.org/MarkUp/>.
- [W3Hc] *A little history of the world wide web*, <http://www.w3.org/History.html>.
- [W3Hd] *World wide web consortium*, <http://www.w3.org/Consortium/>.
- [W3M] *The semantic web made easy*, <http://www.w3.org/RDF/Metalog/docs/sw-easy>.
- [W3O] *Web-ontology (webont) working group*, <http://www.w3.org/2001/sw/WebOnt/>.
- [W3R] *Resource description framework (rdf)*, <http://www.w3.org/RDF/>.

- [W3S] *Semantic web*, <http://www.w3.org/2001/sw/>.
- [W3U] *Naming and addressing: Uris, urls, ...*, <http://www.w3.org/Addressing/>.
- [W3W] *Web services activity*, <http://www.w3.org/2002/ws/>.
- [W3Xa] *Extensible markup language (xml)*, <http://www.w3.org/XML/>.
- [W3Xb] *Xml schema*, <http://www.w3.org/XML/Schema>.
- [Wika] *Teoria grafów*, [http://pl.wikipedia.org/wiki/Teoria\\_graf%C3%B3w](http://pl.wikipedia.org/wiki/Teoria_graf%C3%B3w).
- [Wikb] *Wikipedia, the free encyclopedia*, <http://www.wikipedia.org/>.
- [Wit] Ludwig Wittgenstein, *Tractatus logico-philosophicus, przekład bogusława wolniewicza, zmodyfikowany przez wojciecha sadego, fragmenty*, <http://www.sady.strona.pl/witt.tlp.fr.htm>.
- [Wit00a] ———, *Dociekania filozoficzne*, przeł. Bogusław Wolniewicz, PWN Warszawa, 2000.
- [Wit00b] ———, *Tractatus logico-philosophicus*, przeł. Bogusław Wolniewicz, PWN Warszawa, 2000.

## Spis rysunków

1	Ilość węzłów–domen podłączonych do internetu w latach 1994–2004 . . . . .	9
2	Przykład struktury publikacji HTML w reprezentacji <i>grafu skierowanego</i> . . . . .	18
3	Dokument HTML „Teoria grafów” w reprezentacji <i>grafu skierowanego</i> . . . . .	19
4	Wyświetlanie przykładowego pliku HTML w przeglądarce WWW. . . . .	22
5	Logo <i>World Wide Web Consortium</i> . . . . .	23
6	Wieża <i>Semantic Web</i> . . . . .	28
7	Wieża <i>Sieci Semantycznej</i> . . . . .	28
8	Logo <i>Resource Description Framework</i> . . . . .	39
9	Diagram relacji między zasobami . . . . .	40