1 CYC-realizacja zdrowego rozsądku

Najbardziej ambitny system oparty na regułach zawartych w bazie wiedzy, realizowany od 1984 roku, znany jest pod akronimem CYC. Douglas Lenat, kierownik projektu, ocenia liczbę reguł systemu posiadającego zdrowy rozsądek na rzędu 100 milionów! Według niego w programach AI brakuje przede wszystkim dostatecznie szerokiej bazy wiedzy. Programy, które używają obszernych baz wiedzy to systemy doradcze lub eksperckie. Pokazały one, że nawet niewielka baza wiedzy, rzędu 100-1000 reguł w jakiejś wąskiej dziedzinie, prowadzić może do interesujących rezultatów. Jednakże programy takie są mało odporne na niewielkie nawet odstępstwa od ścisłości sformułowań przy zadawaniu pytań. Niespodziewane sytuacje łatwo załamują systemy eksperckie. Ludzie tak łatwo nie dają się zbić z tropu. Dlaczego? Ponieważ mogą odpowiadać na wiele różnych sposobów, mogą „wyjść z sytuacji". Porządna baza wiedzy powinna zawierać nie tylko ogólne fakty i heurystyki ale i wiele specyficznej wiedzy w wielu dziedzinach. Powinna zawierać informację o nastawieniach różnych ludzi i ich światopoglądzie, różne sposoby patrzenia na rzeczy, mikroteorie zależne od kontekstu.

Oprócz „kruchości" systemów AI motywacją do rozwoju systemu CYC była chęć sprawdzenia, czy istniejące sposoby reprezentacji wiedzy są wystarczające, czy stanowią odpowiednio dobre przybliżenie do zdroworozsądkowej wiedzy. Reprezentacja podstawowych pojęć, czyli podstawowa ontologia dotycząca czasu i przestrzeni, ruchu, substancji, powinna być wspólna dla różnych systemów ekspertowych, jednakże nikt takiej reprezentacji nie stworzył. systemy oparte na wiedzy (knowledge-based systems) muszą mieć wspólną podstawę porozumiewania się. Zdecydowano zakodować ogromną bazę wiedzy w postaci reguł, nie posługując się technikami uczenia maszynowego a jedynie analizą dokonywaną przez ludzi. CYC jest pierwszym projekt tego typu. Stworzenie takiej dużej bazy danych pozwoli w pewnym momencie przekroczyć barierę gromadzenia nowej wiedzy, którą jest niejako „ręczne" wprowadzanie wiedzy, a zastosować bardziej automatyczne, w której system analizy języka naturalnego na tej bazie wiedzy oparty będzie sam mógł wydobywać z tekstu fakty, których jeszcze nie zna. Wymaga to nie tylko bazy wiedzy ale i języka, w którym reprezentuje się wiedzę deklaratywną i procedur manipulowania tą wiedzą.

CYC oparty jest na specyficznym sposobie opisywania (reprezentacji) świata, pozwalającym na uwzględnienia nie tylko obiektów, lecz i takich pojęć jak zdarzenia, nastawienia. Język reprezentacji wiedzy CYCL w tym projekcie użyty rozwijał się stopniowo wraz z samą bazą wiedzy. Jest on oparty na ramach, dla których zdefiniowano rachunek predykatów wraz z możliwościami uzupełniania zmiennych domyślnych. Wprowadzanie nowych pojęć od czasu do czasu prowadzi do konieczności rozszerzenia tego języka, jeśli trudno jest w nim coś wyrazić. Ta „łatanina" była co jakiś czas wygładzana i uzgadniana z poprzednio zaakceptowanymi rozszerzeniami. Po dłuższym okresie czasu okazało się, że język w znacznej mierze się ustabilizował i nie pojawiały się już nowe pojęcia wymagające jego modyfikacji.

Ramy często zawierają mechanizmy dziedziczenia tylko dla specjalnych połączeń, np. isa lub instance. Zdanie „wszystkie ptaki maja dwie nogi" można zakodować ustalając domyślną liczbę nóg w szufladce „nogi" dla wszystkim ramek typu „ptak". Jeśli mamy zdanie „wszyscy przyjaciele królika lubią marchewkę" to każda ramka z „królik" w szufladce „przyjaciel" powinna automatycznie odziedziczyć „marchewka" jako zawartość szufladki „lubi". Dziedziczenie odbywać się może również przez cały łańcuch relacji, a więc stwierdzenia typu „wszyscy krewni przyjaciół królika mają długie uszy" wystarcza by zapełnić szufladki „uszy" krewnych przyjaciół królika informacją „długie". Oprócz ram CYCL zawiera możliwości opisu poprzez specyfikację ograniczeń (constraint language), np. „Jan lubi takich ludzi, którzy potrafią programować w Fortranie" nie służy do zapełnienia szufladek wszystkich ludzi stwierdzeniem „Jan ich lubi" ale jest zmienną ograniczająca fortran-constraint, przypisaną szufladce „lubi" ramy Jana. Ten język specyfikacji ograniczeń pozwala na realizację pełnej logiki pierwszego rzędu, a więc stwarza większe możliwości niż język ram. Jest on traktowany jako uzupełnienie ze względu na większą sprawność rozumowania opartego o ramy niż rozumowania wymagającego logicznych wniosków.

Mechanizmy wnioskowania również rozwijały się krok po kroku. W miarę identyfikacji często używanych klas wnioskowań wprowadzano mechanizmy ich sprawnego przeprowadzenia. W ten sposób otrzymano szereg specyficznych mechanizmów wnioskowania, zależnie od dziedziny. Tradycyjne podejście poszukiwało jakiegoś uniwersalnego mechanizmu dla rozwiązywania problemów. Wnioskowanie wiąże się z tworzeniem „mikroteorii", czyli opisu posługiwania się takimi rzeczami jak pieniądze, robienie zakupów, prowadzenia pojazdu itp. By stwierdzić, czy system rozumie dostatecznie dobrze daną dziedzinę daje mu się różne

opowiadania do analizy i zadaje pytania, na które każdy człowiek powinien umieć odpowiedzieć. CYC zawiera ponad 20 różnych mechanizmów wnioskowania i dla każdego z nich usiłuje zachować spójność wiedzy. Wymaga to określenia, czy w szufladkach są dopuszczalne wartości oraz mechanizmu rozstrzygania sprzeczności.

Czas, w którym stosowane jest domyślne rozumowanie, zależy od reguły: część reguł traktowana jest jako reguły używane tylko wtedy, gdy się do nich odwołujemy (if-needed rule), pozostałe to reguły oceniające wszystko, co się da (if-added rule). Np. jeśli dodamy do własności znajomego Jana „programuje w Fortranie" to nazwisko tego znajomego może się pojawić na liście osób, które Jan lubi. Takie reguły (nazywane też forward rules, czyli wyprzedzające) wymagają dużo czasu i generują sporo mało przydatnych faktów. By nie zabierać czasu zwykle podsystem utrzymywania spójności CYC (truth maintenance system) dodaje nowe fakty w nocy lub w okresie przerwy w jego wykorzystaniu. Język specyfikacji ograniczeń pozwala na utworzenie abstrakcyjnego poziomu reprezentacji, określanej jako poziom epistemologiczny (EL). Poziom reprezentacji heurystycznej odnosi się do faktów korzystających z bardziej efektywnych, heurystycznych sposobów rozumowania, podczas gdy poziom EL wymaga abstrakcyjnego rozumowania logicznego.

Z każdą szufladką skojarzony jest zbiór mechanizmów wnioskowania, jakim można się posłużyć w celu zapełnienia jej wartości. W najprostszym przypadku zapytanie skierowane do systemu wymaga sprawdzenia zawartości szufladki, w bardziej złożonych przypadkach wymagane jest uruchomienie mechanizmów wnioskujących w celu zapełnienia szufladki, a w miarę wzrostu komplikacji pytań konieczna staje się strategia kontrolna, którą również zaimplementowano przy pomocy ram. Schematy postępowania czy wnioskowania, zawarte w ramach, można kopiować i modyfikować dostosowując je do nowych przypadków, a kompilator języka CYCL tworzy kod w Lispie. Chociaż CYC oparty jest na wiedzy deklaratywnej, to jego mechanizmy wnioskowania pozwalają na realizacje pewnych procedur. Jego autorzy podkreślają, że CYCL nie jest typowym językiem reprezentacji wiedzy opartym na ramach, a raczej wykorzystuje logiczne stwierdzenia (assertions). CYC zawiera około pół miliona istotnych stwierdzeń, prostych faktów, reguł i sposobów wnioskowania (już w 1987 roku jego baza przekroczyła 100 tysięcy faktów); każde stwierdzenie może być powiązane z kilkudziesięcioma innymi, stąd liczba powiązań w tym systemie sięga milionów. Pierwsze ponad milion reguł/powiązań w systemie CYC odnosiło się nie tyle do szczegółowych faktów co do spraw ogólnych, klasyfikacji, ograniczeń, takich pojęć jak czas, przestrzeń, substancja, czyli do globalnej ontologii. Pojęcia takie trudno jest reprezentować w bazie wiedzy - rozwiązanie w tym systemie polega na opisie lub relacjach, które sprawdzają się bardzo dobrze w najczęstszych przypadkach użycia danej koncepcji i które dobrze działają również w dość często spotykanych sytuacjach. Zamiast próbować zredukować, znaleźć minimalną liczbę sposobów oddziaływania podmiotów i przedmiotów próbuje się opisać wszystkie sytuacje.

CYC jest pierwszym programem w którym dyskutuje się zagadnienia globalnej ontologii czyli klasyfikacji bytów. Na najwyższym poziomie ma on koncepcję Rzeczy, która może być rzeczą konkretną lub abstrakcyjną. Wszystko jest rodzajem rzeczy. Ważnym rozróżnieniem są rzeczy indywidualne i kolekcje, czyli zbiory rzeczy. Ramy, które należą do kolekcji rzeczy, to osoba, naród, nos. Rzeczy indywidualne to Jan, Polska, nos Jana. Rzeczy indywidualne mogą mieć części a kolekcje mogą mieć zbiory i podzbiory. Rzeczy nienamacalne nie maja masy, są to zdarzenia, liczby, prawa. Rzeczy namacalne mają masę, np. ciało człowieka, jabłko czy kurz. Rzeczy złożone mają zarówno cechy namacalne jak i nienamacalne, np. osoba ma ciało i umysł. Substancja jest przykładem ObiektuIndywidualnego. Substancja zachowuje swoje własności jeśli jest pocięta na mniejsze kawałki. Własności mogą być zewnętrzne (extrinsic) jak i wewnętrzne (intrinsic), wtedy gdy własność odnosi się zarówno do części obiektu jak i całego obiektu. Zdarzenia to rzeczy dziejące się w czasie, a procesy to szczególny rodzaj zdarzeń, które definiuje się podobnie jak substancję - po podzieleniu mają zachowywać własności. Spacer jest procesem, ale kilometrowy-spacer nie jest. Zdarzenia mają własności temporalne, interwały czasowe i zbiory interwałów. Slot, czyli szufladka, to podklasa nienamacalnego. Jest wiele rodzajów szufladek: definiujące, buchalteryjne (zapisujące informacje o pochodzeniu), ilościowe itp. Agent, rodzaj obiektu złożonego, to zbiór inteligentnych istot. Mogą to być ludzie czy firmy. Agenci mają przekonania (beliefs) i mogą przypisywać przekonania innym. Ponieważ te przekonania nie zawsze są prawdziwe CYC musi odróżnić lokalne przekonania agentów od swojej własnej wiedzy o świecie.

Przykład wiedzy systemu CYC na temat kupowania:

1. Pieniądze daje się w zamian za towary, rzeczy, usługi lub jako dar.
2. Każdy czynnik jest związany z pewną sumą pieniędzy.
3. Opłaty do 10$ dokonywane są zwykle gotówką, powyżej 50$ czekiem lub kartą kredytową.
.....