

Architektury kognitywne, czyli jak zbudować sztuczny umysł.

Włodzisław Duch

Katedra Informatyki Stosowanej, UMK, Toruń

Streszczenie.

Architektury kognitywne (AK) są próbą stworzenia modeli komputerowych integrujących wiedzę o działaniu umysłu. Ich zadaniem jest implementacja konkretnych schematów działania funkcji poznawczych umożliwiającą testowanie tych funkcji na szerokiej gamie zagadnień. Wiele architektur kognitywnych opracowano w celu symulacji procesu komunikacji pomiędzy człowiekiem i złożonymi maszynami (HCI, Human-Computer Interfaces), symulowania czasów reakcji oraz różnych psychofizycznych zależności. Można to do pewnego stopnia osiągnąć budując modele układu poznawczego na poziomie symbolicznym, z wiedzą w postaci reguł logicznych. Istnieją też projekty, które próbują powiązać procesy poznawcze z aktywacją modułów reprezentujących konkretne obszary mózgu, zgodnie z obserwacjami w eksperymentach z funkcjonalnym rezonansem magnetycznym (fMRI). Dużą grupę stanowią architektury oparte na podejściu logicznym, które mają na celu symulację wyższych czynności poznawczych, przede wszystkim procesów myślenia i rozumowania. Niektóre z projektów rozwoju architektur poznawczych skupiają większe grupy badawcze działające od wielu dziesięcioleci.

Ogólnie architektury kognitywne podzielić można na 3 duże grupy: architektury symboliczne (oparte na funkcjonalnym rozumieniu procesów poznawczych); architektury emergentne, oparte na modelach koneksjonistycznych; oraz architektury hybrydowe, wykorzystujące zarówno modele neuronowe jak i reguły symboliczne. W

ostatnich latach znacznie wzrosło zainteresowanie architekturami inspirowanymi przez neurobiologię (BICA, *Brain Inspired Cognitive Architectures*). Jak sklasyfikować różne architektury, jakie wyzwania należy przed nimi postawić, jak oceniać postępy w ich rozwoju, czego nam brakuje do stworzenia pełnego modelu umysłu? Krytyczny przegląd istniejących architektur kognitywnych, ich ograniczeń i możliwości pozwala na sformułowanie ogólnych wniosków dotyczących kierunków ich rozwoju oraz wysunięcie własnych propozycji budowy nowej architektury.

1. Wstęp

Długofalowym zadaniem stojącym przed badaniami nad sztuczną inteligencją jest stworzenie systemów, które osiągną poziom kompetencji przekraczający możliwości ludzkie w różnych dziedzinach. W ograniczonym zakresie jest to już oczywiście możliwe: komputery szybciej liczą, lepiej pamiętają, analizują modele zbyt złożone dla człowieka z powodu złożoności procesów, dużej ilości danych lub liczby zmiennych poddawanych optymalizacji. Komputery dokładniej rozpoznają specyficzne wzorce w sygnałach, lepiej sobie radzą niż ludzie w wielu grach planszowych, prowadzą manewry sterując helikopterami, samolotami i pojazdami naziemnymi w sposób niemożliwy dla ludzi. Jednakże pomimo tego ciągłego postępu w wielu dziedzinach sztuczna inteligencja całkowicie zawiodła pokładane w niej nadzieje. Stosunkowo proste niższe funkcje poznawcze, związane z percepcją, rozpoznawaniem obiektów i analizą relacji pomiędzy nimi, nadal pozostają daleko poza możliwościami nawet prymitywnych gatunków zwierząt. Znaczny postęp nastąpił jedynie na poziomie stosunkowo prostej analizy wzorców i przetwarzania sygnałów. Wyższe czynności poznawcze, związane z użyciem rozpoznanych wzorców – słów, symboli, obiektów – do wnioskowania, analizy tekstów, dialogu w języku naturalnym, planowania i rozwiązywania problemów, tworzenia i wykorzystywania reprezentacji złożonych form wiedzy, okazały się znacznie trudniejsze do symulacji. Do takich zastosowań nie wystarczą proste inspiracje biologiczne dotyczące działania pojedynczych neuronów lub ich niewielkich grup. Konieczne jest uwzględnienie wielkoskalowej architektury mózgu, modele funkcji wyspecjalizowanych obszarów, które coraz lepiej uczymy się poznawać.

Takie podejście można określić mianem **informatyki neurokognitywnej**. Funkcje poznawcze wspomagane są przez pamięć różnego typu:

- pamięć rozpoznawczą, pozwalającą na identyfikację znanych obiektów, lub dostrzeżenie odstępstw od oczekiwań;
- pamięć skojarzeniową, prowadzącą automatycznie do prostych wniosków, realizację procesów warunkowania klasycznego;
- pamięć proceduralną, czyli pamięć umiejętności manualnych i sekwencji działań;
- pamięć semantyczną, pozwalającą na interpretację sensu i dostęp do złożonych struktur wiedzy;
- pamięć roboczą, pozwalającą na łączenie ze sobą w kombinatoryczny sposób różnych fragmentów informacji w większe całości.

Wszystkie te rodzaje pamięci wspierane są przez sieci neuronowe o specyficznej architekturze, dostosowanej do wymaganych funkcji (O'Reilly, Munakata, 2000). Organizacja przechowywania i dostępu do informacji przez mózgi i komputery różni się całkowicie na bardzo podstawowym poziomie. Matematyczne dowody uniwersalności komputerów (maszyny Turinga, sieci neuronowych) do pewnego stopnia zaślepiły badaczy – znane nam architektury obliczeniowe ograniczają na wiele sposobów przetwarzanie informacji. W niektórych zastosowaniach komputery robią to lepiej niż mózgi, ale nadal w wielu ważnych zastosowaniach mózgi są niedoścignionym wzorem. Nie wiemy, czy architektury kognitywne zaimplementowane na konwencjonalnych komputerach mogą osiągnąć podobne możliwości, zarówno w przypadku niższych jak i wyższych czynności poznawczych.

Sztuczna inteligencja (*artificial intelligence*, AI) zajmowała się tradycyjnie wyższymi czynnościami poznawczymi: rozwiązywaniem problemów, myśleniem, reprezentacją wiedzy, analizą języka naturalnego (Russell i Norvig, 2003). Próbowano tego dokonać w oderwaniu od mechanizmów biologicznych, w oparciu o symboliczne modele złożonych form wiedzy, a więc niejako na poziomie umysłu, wiedzy dającej się poznać w świadomy sposób. Cechą charakterystyczną takiego podejścia jest skupienie się na

procesach sekwencyjnych, w których złożoność kombinatoryczna, wynikająca z możliwości zastosowania różnych transformacji przydatnych w rozwiązywaniu problemu, wymaga rozważenia bardzo wielu możliwości i wyboru odpowiedniej strategii szukania rozwiązania. Podstawę dla działań inteligentnych stanowią w tym ujęciu algorytmy szukania rozwiązań i symbolicznej reprezentacji wiedzy (Newell, 1990). Niższe czynności poznawcze, realizowane w znacznie krótszej skali czasowej, w całkowicie równoległy sposób, umożliwiają percepcję, postrzeganie obiektów, kategoryzację, szybkie skojarzenia, działania sensomotoryczne. Takie czynności są podstawą działania robotów. Tendencje rozwojowe w tej dziedzinie, zmierzające do stworzenia inteligentnych robotów, które będą się rozwijać podobnie jak dzieci, ugruntowując sens symboli i realizowanych poleceń w działaniu w świecie, doprowadziły do upowszechnienia się dominującej obecnie filozofii „ucieleśnienia poznania” (*embodied cognition*) oraz enaktywizmu (Barsalou, 2008), zgodnie z którą sensem symboli są specyficzne sekwencje działań sensomotorycznych. Jest tu wiele problemów, szczególnie ze zrozumieniem w jaki sposób tworzą się abstrakcyjne reprezentacje mentalne w oparciu o reprezentacje percepcyjno-ruchowe (Mahon i Caramazza, 2008). Symboliczne aproksymacje działań sensomotorycznych mają bez wątpienia silne ograniczenia, podobnie jak wszelkie próby dyskretyzacji ciągłych procesów. Nie wiemy, jak daleko trzeba pójść w stronę ucieleśnionych reprezentacji podstawowych procesów by sens, nadany pierwotnym koncepcjom przez działanie w świecie był wystarczający do stworzenia na tej podstawie bardziej abstrakcyjnych koncepcji (pisałem o tym już w Duch, 1994, 1996). Być może jest jeszcze za wcześnie by porzucać marzenia Newella i Simona (1963) o konstrukcji Ogólnego Rozwiązywacza Problemów (General Problem Solver, GPS) na poziomie manipulacji symbolami, chociaż nie da się zapewne uniknąć rozszerzeń reprezentacji symbolicznych w stronę lepszej aproksymacji procesów skojarzeniowych. Zastosowania w robotyce, sterowaniu, czy próby zrozumienia zachowania organizmów biologicznych mogą wymagać dokładniejszych modeli, uwzględniających ucieleśnienie poznania na głębszym poziomie.

Wiele ciekawych idei działających dla prostych problemów nie sprawdza się dobrze w rzeczywistych, złożonych zastosowaniach. Pojawiła się ostatnio tendencja do budowy ogólnej sztucznej inteligencji (Artificial General Intelligence, AGI), a więc systemów, które nie tylko potrafią rozwiązać jedno zadanie na poziomie mistrzowskim (np. grać w szachy lepiej od mistrza świata), lecz dają się zastosować do całej klasy podobnych problemów (np. nauczyć się innych gier i osiągnąć w nich wysoki poziom). Lista wyzwań, stojących przed inteligencją obliczeniową jest długa (Duch i Mandziuk, 2007). Analiza istniejących architektur poznawczych powinna pomóc w zrozumieniu ich ograniczeń i możliwych sposobów ich przewyżnienia.

2. Wielkie wyzwania

Czego należy oczekiwać od architektur poznawczych, by można je było uznać za coś więcej niż wyspecjalizowane programy udające inteligencję? AI skupiło się nad praktycznymi problemami tworzenia systemów doradczych, prawie całkowicie porzucając swoje pierwotne ambitne cele, takie jak stworzenie GPS (Newell i Simon, 1963). Okazało się, że nie wystarczy do tego wiedza ogólna, rozwiązywanie problemów wymaga bardzo obszernej wiedzy w wielu dziedzinach. Projekt Cyc Douglasa Lenata stworzenia systemu zawierającego encyklopedyczną wiedzę (nazwa „Cyc” jest fragmentem ang. enCYClopedia), konieczną do realizacji zdrowego rozsądku, jest jak dotychczas jedyną taką próbą (Panton i inn, 2006), realizowaną w oparciu o klasyczne metody AI. Projekt ten rozpoczęto w 1984 roku, konstruując ogromną bazę wiedzy w oparciu o złożone struktury, nazywane ramami. Ontologia systemu Cyc zawiera ponad 300 tysięcy pojęć i wiele milionów łączących je faktów (OpenCyc w wersji 1.0 miał ponad 3 miliony asercji dotyczących tych pojęć, częściowo ograniczających i definiujących wzajemne relacje ponad 26 tysięcy typów), a w jego rozwój włożono ponad 600 osobolat pracy. Rezultaty nie są zachwycające i nadal brakuje ambitnych zastosowań, które pokazałyby, że tak złożony system istotnie daje się do czegoś użyć, chociaż zawarta w niej wiedza jest ostatnio wykorzystywana przez projekty internetu semantycznego. Lista „potencjalnych zastosowań” Cyc jest długa, ale tylko CycSecure,

program do analizy zabezpieczeń sieci komputerowych, jest realnym zastosowaniem, prezentowanym na oficjalnej stronie www.cyc.com firmy Cycorp.

AI zawiodła w wielu dziedzinach, ale rzeczywistość chyba najbardziej odbiega od oczekiwań w dziedzinie systemów do dialogu w języku naturalnym i innych zagadnieniach dotyczących rozumienia języka. Napisano wiele programów do konwersacji, które nazywają się botami lub chatterbotami (od nazwy ro-bot powstały nazwy typu soft-bot, skrócone do bot). Programy te znajdują zastosowanie w firmach komercyjnych do odpowiadania na pytania dotyczące działalności firmy. Test Turinga (1950), polegający na konwersacji z ludźmi i botami na niczym nieograniczony temat, jest częściowo stosowany w konkursach Loebnera (<http://www.loebner.net>). Od prawie 20 lat postęp jest tu jednak niewielki, w większości przypadków stosuje się techniki oparte na dopasowaniu odpowiedzi do szablonów pytań (Wallace, 2003), znane od ponad pół wieku. Programy oparte na szablonach nie mają pojęcia, jaki jest sens koncepcji pojawiających się w dialogu, nie mają żadnego strukturalnego opisu pojęć. Dlatego nie mogą rozwinąć rzeczywistego zrozumienia sensu tekstów i dialogów, można je co najwyżej użyć do stereotypowej konwersacji w systemach odpowiadających na pytania w wąskiej dziedzinie. Niektórzy sędziowie w konkursie Loebnera dają się oszukać programom przejmującym inicjatywę prowadzenia konwersacji, zadającym dużo pytań, dzięki czemu łatwiej jest im zawęzić oczekiwania i analizować odpowiedzi dyskutantów za pomocą szablonów. Proponowano liczne zmodyfikowane wersje testu Turinga, np. Carpenter i Freeman (2005) piszą o „spersonalizowanym teście”, w którym program ma się wcielić w osobę znaną rozmówcy, jest to więc próba określenia na ile można zrobić model określonego umysłu, a nie jakiś ogólny, abstrakcyjny program do konwersacji.

Ciekawe wyzwania dla AI stwarzają gry słowne, a w szczególności gra w 20 pytań, stanowiąca interesujący paradygmat dla wielu innych zastosowań. Gry słowne wymagają szerokiej wiedzy o znaczeniu słów, oraz o własnościach obiektów wskazywanych przez te słowa, ale nie wymagają pełnej wiedzy o złożonych relacjach pomiędzy tymi obiektami. W takich zastosowaniach okazuje się, że nie ma jednej,

uniwersalnej i optymalnej metody reprezentacji wiedzy. Złożone formy wiedzy, np. ramy Cyc, trudno jest szybko przeszukiwać jeśli chcemy określić, jakie pytanie warto zadać by zdobyć maksymalnie dużo informacji. Proste wektorowe metody reprezentacji informacji mogą w tym przypadku być bardziej przydatne (Szymański i inn. 2007).

Jednakże nadal potrzebna jest obszerna wiedza o tysiącach różnych koncepcji, a niełatwo jest ją automatycznie stworzyć, pomimo tego, że dostępnych jest sporo ontologii, słowników, encyklopedii i słabiej ustrukturalizowanych źródeł wiedzy.

Systemy do odpowiadania na pytania (Q/A systems) są jeszcze bardziej wymagające, gdyż zrozumienie pytania wykracza poza własności koncepcji i wymaga zrozumienia sensu zdania. Seria konferencji TREC (Text Retrieval Conference)¹ organizuje konkursy dla systemów wyszukiwawczych, które mają znaleźć w Internecie lub w swoich bazach wiedzy odpowiedzi na zbiór pytań konkursowych. Innym ważnym wyzwaniem są inteligentne systemy wspomagające nauczanie (Intelligent Tutoring Systems, ITS), chociaż nie ma w tej dziedzinie jednoznacznej metody na określenie postępów.

Budowa systemów doradczych, które mogłyby służyć jako partnerzy dla ludzi, jest nadal wielkim wyzwaniem (Feigenbaum, 2003). Chodzi tu nie tyle o klasyczne systemy doradcze, co o systemy obdarzone szerszą inteligencją, pozwalające na łatwiejszą komunikację z człowiekiem, rozumujące i podpowiadające interesujące skojarzenia w takich dziedzinach jak matematyka, nauki o życiu czy prawo. Rezultaty łatwo jest ocenić: panel ekspertów może przeprowadzić egzamin, zadając pytania i domagając się szczegółowego uzasadnienia, w celu oceny stopnia zrozumienia danego problemu.

Potrzebne są tu nie tylko modele prowadzące ścisłe wnioskowanie logiczne, ale i modele skojarzeniowe, zauważające interesujące korelacje i powiązania. Konkursy dowodzenia twierdzeń matematycznych odbywają się na konferencjach CADE (Conference on Automated Deduction)² w wielu specjalistycznych kategoriach. Rozszerzenie możliwości takich wyspecjalizowanych modułów na meta-poziom pozwalający na

¹ <http://trec.nist.gov/>

² <http://www.cs.miami.edu/~tptp/CASC/>

komunikację z użytkownikiem i wybór odpowiedniego modułu do danego problemu jest drogą do stworzenia interesującego partnera dla matematyka. Złożoność genomicznych, metabolicznych i proteomicznych procesów jest zbyt duża, by człowiek mógł je poznać i zrozumieć, dlatego super-ekspert w tej dziedzinie musi powstać w bliskiej przyszłości.

Podobne wyzwania warto formułować w innych dziedzinach. Odległym celem jest tu stworzenie komputerowych doradców dla ekspertów – programów oceniających ich rozumowanie, wspomagających to rozumowanie, podsuwających interesujące obserwacje i skojarzenia, a nawet kreatywne idee (Duch i Pilichowski, 2007).

Amerykańska agencja departamentu obrony do zaawansowanych projektów badawczych (Defense Advanced Research Projects Agency, DARPA)³ sponsoruje zakrojony na szeroką skalę program budowy spersonalizowanych asystentów osobistych (Personalized Assistants that Learn, PAL, z budżetem ponad 7 milionów \$), i innych systemów kongitywnych⁴.

Miarą ogólnej inteligencji systemu AI może być liczba zadań, lub nawet całych profesji, które dany program może w pełni zautomatyzować (Nilsson, 2005). Już Allan Turing zauważył (1950), że taki ogólny system (nazwał go *child machine*), który mógłby nauczyć się różnych zadań, wystarczy by zastąpić wiele specjalizowanych systemów. Wiele zawodów związanych jest obecnie z przetwarzaniem informacji i przydatność programów można w nich mierzyć za pomocą standardowych testów egzaminacyjnych. Zawody związane z wykonywaniem fizycznej pracy, wymagające koordynacji sensomotorycznej i rozwiązania problemów związanych z percepcją (w tym analizą obrazów), wymagać będą autonomicznych robotów. DARPA zorganizowała dwa udane konkursy dla automatycznie sterowanych pojazdów, przejazdu przez pustynię i jazdy w symulowanych warunkach miejskich (Darpa Urban Challenge Competition 2007). Takie zadanie wymagało przede wszystkim integracji metod przetwarzania sygnałów, analizy obrazu, sterowania, planowania i rozumowania w jednym systemie. Jazda samochodem

³ <http://www.darpa.mil/>

⁴ http://www.darpa.mil/ipto/thrust_areas/thrust_cs.asp

to zadanie znacznie prostsze niż kontrola humanoidalnego domowego robota, który miałby współpracować z ludźmi, a więc musiałby rozumieć ich intencje, rozpoznawać obiekty i relacje między nimi, kontrolować uwagę, oraz uczyć się na podstawie obserwacji. Jednakże nie każdy człowiek może się nauczyć jazdy samochodem. W tym zakresie automatycznie sterowane pojazdy osiągnęły więc wyższy poziom inteligencji niż niektórzy ludzie. Firmy samochodowe zapowiedziały w ciągu najbliższego dziesięciolecia zastąpienie zawodowych kierowców na dalekich trasach przez systemy automatyczne⁵.

Stworzenie osobistych asystentów nie musi w pełni eliminować pracy człowieka, na początek powinno wspomagać jego produktywność. Brakuje tu uporządkowania stopnia trudności różnych zadań tak, by stawiać sobie realistyczne cele jak i oceny kompetencji takich systemów. Wiele zawodów już zniknęło, nastąpiła znaczna redukcja liczby pracowników związanych bezpośrednio z produkcją jak i zatrudnionych w instytucjach finansowych. W większości przypadków zmniejszenie zatrudnienia to wynik zmiany organizacji pracy, wprowadzenia nowych urządzeń o stosunkowo niskim stopniu inteligencji (automatyka przemysłowa, automaty ATM zamiast kas). Nie ulega jednak wątpliwości, że większość ambitnych projektów w sztucznej inteligencji wymagać będzie systemów, które powinny współdziałać w szerokim zakresie z naszymi umysłami, mając podobne reakcje i funkcje. Takie systemy określane są mianem „architektur kognitywnych”.

3. Architektury kognitywne

Zrozumienie sposobu działania naszego aparatu poznawczego wymaga modelu działania człowieka, który uwzględnia jego możliwości percepcyjne i poznawcze. Początkowo architektury kognitywne tworzone przede wszystkim z myślą o tym, by modelować reakcje człowieka w złożonych sytuacjach, gdy trzeba reagować w odpowiedni sposób na wiele sygnałów o różnych modalnościach (Newell 1990, Meyer i Kieras 1997). Niewłaściwie zaprojektowany system sterowania dużym samolotem może stawiać zbyt

⁵ http://en.wikipedia.org/wiki/Driverless_car

wielkie wymagania w stosunku do możliwości reakcji przez pilotów. Dlatego znajomość ludzkich ograniczeń, jak też wierne modele sposobu ich działania, są bardzo pożądane.

Jak oceniać architektury kognitywne? Allen Newell (1990) zaproponował aż 12 kryteriów oceny takich systemów, oceniając ich zdolności adaptacyjne, zachowania dynamiczne, elastyczność zachowań, możliwości stopniowych ulepszeń, możliwości całkowitej ewolucji architektury, sposoby uczenia się, sposoby integracji wiedzy, wielkość dostępnej bazy wiedzy, kompetencje językowe, działanie w czasie rzeczywistym oraz możliwości implementacji w mózgo-podobnych architekturach. Anderson i Labiere (2003) zastosowali te kryteria do analizy trzech znanych architektur (ACT-R, SOAR i klasycznej architektury koneksjonistycznej). Takie szczegółowe porównania trudno jest zastosować do większej liczby systemów. Zaproponowaliśmy prostszą taksonomię (Duch i inn. 2008), przedstawioną poniżej łącznie z przykładami obecnie rozwijanych architektur kognitywnych. Większość prac w tej dziedzinie koncentruje się na konkretnych architekturach i brak jest ogólnego przeglądu, który dałby szerszą orientację. Wyjątkiem jest artykuł Vernona i inn. (2007), koncentrujący się na możliwościach autonomicznego rozwoju zdolności mentalnych agentów programowych.

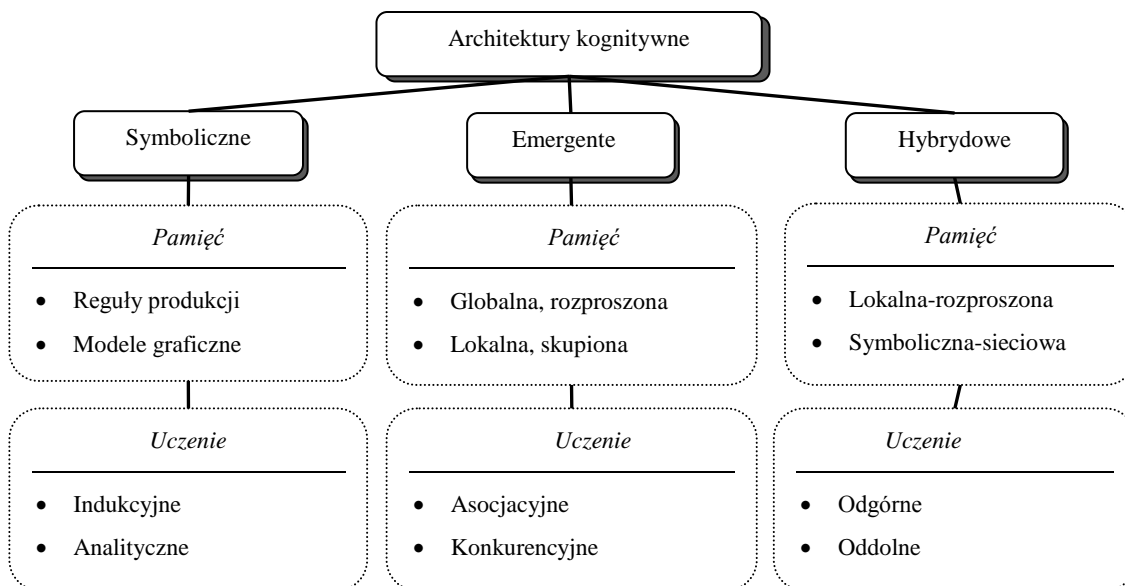
Dwie najważniejsze cechy wszystkich architektur kognitywnych to sposób organizacji ich pamięci i mechanizmy uczenia się. Pamięć jest repozytorium wiedzy o świecie i o sobie, celach i bieżących działaniach. Rola pamięci rozumiana jest w różny sposób przez różnych autorów (Hawkins i Blakeslee 2004, Hoya 2005, Hecht-Nielsen 2007). Uczenie się jest procesem, który transformuje zapamiętaną wiedzę i sposób jej wykorzystania. Pamięć i uczenie się to dwa fundamenty do budowy podstaw systemów poznawczych, na których wyrosnąć mogą bardziej złożone wyższe czynności poznawcze. Organizacja pamięci zależy od sposobu reprezentacji wiedzy. Prosta taksonomia architektur kognitywnych oparta na tych dwóch filarach prowadzi do identyfikacji trzech głównych grup architektur: symbolicznych, emergentnych i hybrydowych.

Architektury symboliczne wykorzystują wiedzę deklaratywną, zawartą w relacjach zapisanych na poziomie symbolicznym, skupiając się nad użyciem tej wiedzy do

rozwiązywania problemów. Architektury emergentne wykorzystują przepływy sygnałów przez sieć licznych, wzajemnie ze sobą oddziałujących elementów, w której pojawiają się stany emergentne, dające się zinterpretować w symboliczny sposób. Architektury hybrydowe są kombinacją obu tych podejść, połączonych na różne sposoby. Typ architektury jest mocno zależny od rodzaju problemów, które zamierza się za jej pomocą rozwiązać. Poniżej przedstawione zostały konkretne przykłady wielu architektur należących do wszystkich opisanych tu kategorii.

3.1 Architektury symboliczne

Hipoteza Newella i Simona (1976) na temat fizycznych systemów symbolicznych ma swoje źródło w badaniach nad pamięcią i rozwiązywaniem problemów. Fizyczny system symboliczny może używać dowolne fizyczne wzorce by stworzyć odpowiednie symbole, manipulować nimi, wczytywać, wypisywać, przechowywać i zmieniać je podejmując przy tym odpowiednie działania ukierunkowane na dążenie do osiągnięcia swoich celów. Chociaż w praktyce idea ta została zawężona do symboli językowych pierwotne sformułowanie jest dość uniwersalne i pozwala na nazywanie symbolami specyficznych konfiguracji pobudzeń układów neuronowych. W takim przypadku



Rys. 1 Uproszczona taksonomia architektur kognitywnych

pojawia się od razu problem dyskretyzacji (zignorowany przez Newella), inaczej liczba symboli oznaczających podobne stany rośnie w sposób kombinatoryczny. Większość architektur symbolicznych używa scentralizowanej kontroli przepływu informacji od czujników sensorycznych do efektorów (kontrola ruchu lub informacji wyjściowych). Można to uznać za aproksymację funkcji wykonawczych związanych z pamięcią roboczą, wspomaganych przez pamięć semantyczną dostarczającą wiedzy. Reguły produkcji mają reprezentować cykl postrzeżenie-działanie, który służy aproksymacji procesów rozumowania przez ekspertów. Stosowane są również reprezentacje grafowe, zwykle w postaci grafów skierowanych, których węzły reprezentują symbole i przypisane im atrybuty, a łuki związki pomiędzy nimi. Tak jest w przypadku sieci semantycznych jak i grafów koncepcyjnych (Sowa, 1984). W graficznej postaci przedstawić też można ramy i schematy (Minsky, 1975). W robotyce stosowana jest reprezentacja za pomocą zbioru reakcji w określonym kontekście (reactive action packages, RAPs; Firby, 1989).

W systemach symbolicznych wprowadzono różne mechanizmy uczenia się, które można podzielić na analityczne (dedukcyjne) i indukcyjne. Uczenie analityczne wykorzystuje wiedzę do tworzenia na jej podstawie nowych faktów. Przykłady tego podejścia to uczenie oparte na wyjaśnieniach (explanation-based learning, EBL, Mitchell i inn. 1986) czy analogiach (Veloso i Carbonell, 1990). Uczenie indukcyjne odkrywa nowe, ogólne reguły na podstawie obserwacji, usiłując dokonać aproksymacji struktury danej domeny. Przykładem jest tu oparte na wiedzy uczenie indukcyjne (knowledge-based inductive learning, KBIL; Larvac i Dzeroski 1994), czy opóźnione uczenie się z krytykiem (Kaelbling i inn. 1996).

Wiele ambitnych architektur symbolicznych przestało rozwijać się po krótkim okresie entuzjazmu ich twórców; poniżej przedstawiono najciekawsze projekty, które rozwijają się od dłuższego czasu lub mają duże szanse na dalszy rozwój.

SOAR (*State, Operator And Result*) jest klasycznym, rozwijanym od ponad 30 lat przykładem architektury kognitywnej służącej modelowaniu ogólnej inteligencji za pomocą systemu regułowego, traktowanego jako przybliżenie do systemów opartych na

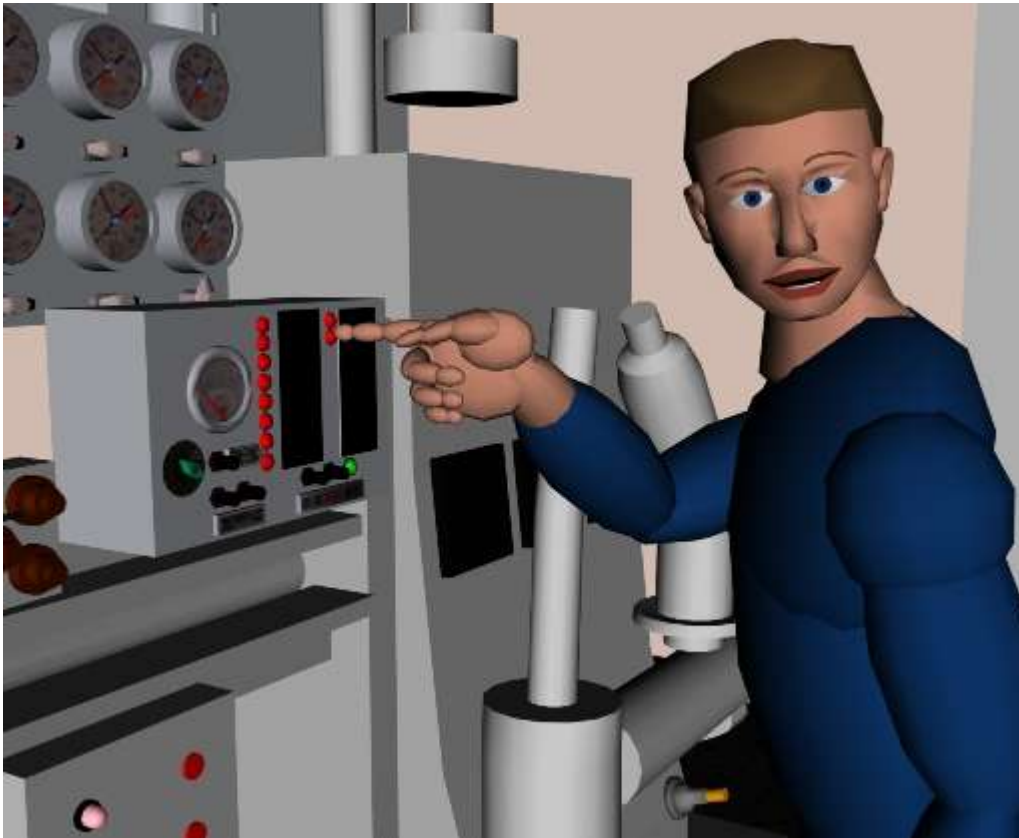
wiedzy (Newell, 1990; Laird i inn. 1987). Sytuacja rozpatrywana jest tu w przestrzeni problemu, czyli zbioru dopuszczalnych stanów opisu problemu. Reguły produkcji “Jeśli spełnione są warunki to należy podjąć działania” (IF ... THEN ...) uporządkowane są według typu operatorów określających działania w tej przestrzeni. Główny mechanizm uczenia oparty jest na idei porcjowania (*chunking*), technice analitycznej formułowania nowych reguł i makro-operacji na podstawie powtarzalnych kombinacji prostych operacji, które okazały się przydatne wcześniej (Laird i inn. 1987). Nowe „porcje wiedzy” prowadzą do rozwiązań niejako na skróty. System dostępny jest w różnych wersjach na stronie projektu⁶.

SOAR zastosowano do replikacji wyników wielu eksperymentów psychologicznych, planowania, pracy z dużymi i złożonymi bazami wiedzy, sterowania eskadrą myśliwców, awatarami w grach komputerowych, rozwiązywania problemów i zrozumienia języka naturalnego (NL-SOAR) w czasie rzeczywistym⁷.

Architektura SOAR uległa w ostatnich latach znacznej ewolucji: uczenie z krytykiem zastosowano do określenia strategii nadawania preferencji operatorom, wprowadzono uczenie epizodyczne do analizy ewolucji stanu problemu, pamięć semantyczną do opisu abstrakcyjnej wiedzy deklaratywnej. Pojawiły się też modne inspiracje biologiczne (Laird, 2008): wyobraźnia przestrzenna w problemach wymagających działania w rzeczywistym świecie, emocje, nastroje i uczucia ukierunkowujące rozumowanie i przyspieszające uczenie się z krytykiem. Proponowane rozszerzenia nie zostały jeszcze w pełni zintegrowane z całą architekturą i nie wiadomo, na ile będzie to możliwe. System percepcyjno-ruchowy zastosowany w SOAR jest dość prymitywny, trzeba w nim definiować własne funkcje określające informację wejściową i wyjściową. Wśród propozycji rozszerzeń brakuje mechanizmów skupiania uwagi, selekcji informacji i zapominania, uczenia hierarchicznych reprezentacji (czyli aproksymacji problemów na różnym poziomie abstrakcji), jak i traktowania niepewności informacji.

⁶ <http://sitemaker.umich.edu/soar/home>

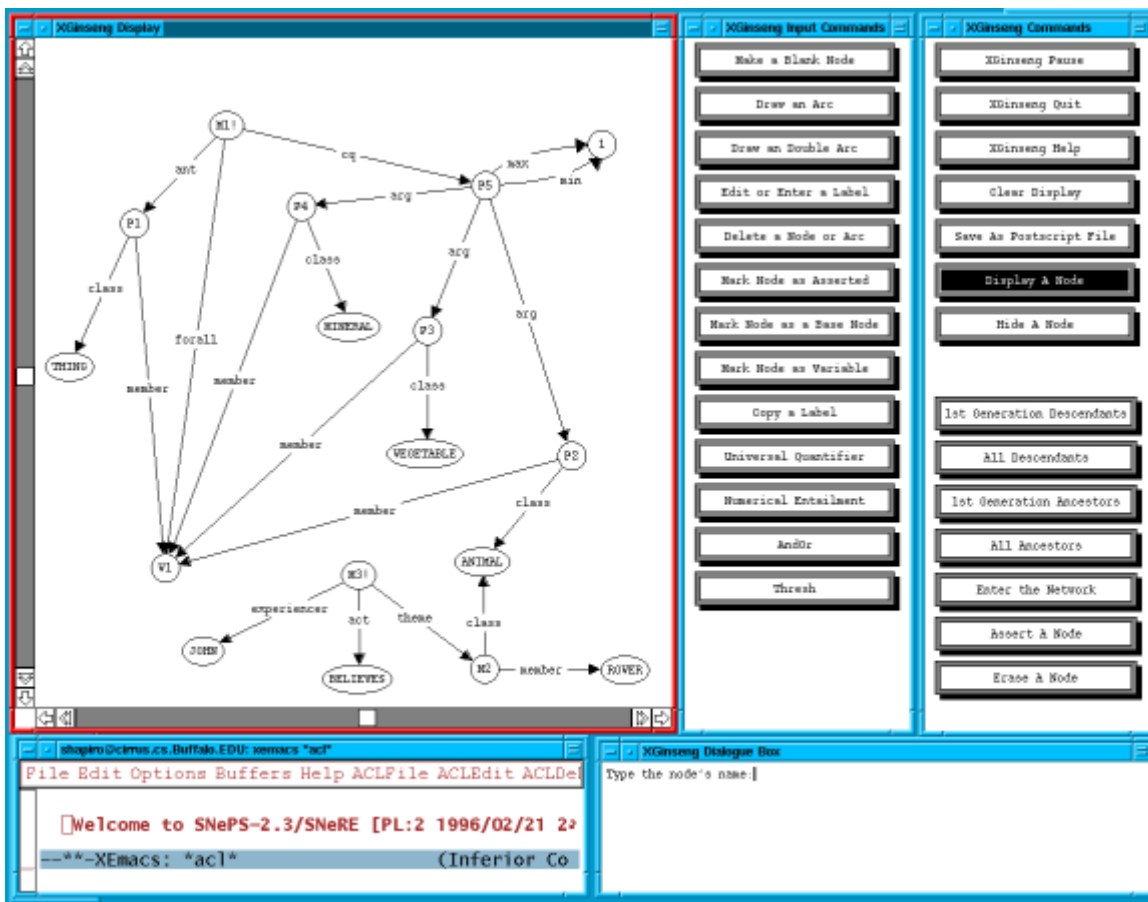
⁷ Opis nowości w SOAR jest pod: <http://ai.eecs.umich.edu/people/laird/current-research.html>



1 Steve, awatar sterowany za pomocą SOAR, pomaga nauczać obsługi skomplikowanej maszyny w wirtualnym świecie.

EPIC (*Executive Process Interactive Control*) to architektura kognitywna, która ma za zadanie modelować wiele aspektów działań człowieka w procesie interakcji z maszynami (Meyer i Kieras, 1997). Kilka połączonych ze sobą procesorów służy do w miarę realistycznego (z punktu widzenia czasów reakcji) przetwarzania informacji przez zmysły wzroku, słuchu i dotyku, a efekty poruszają kamerami, wciskają klawisze i produkują słowne wypowiedzi. Procesory działają na sygnałach przetworzonych do postaci symbolicznej, wykorzystywanej przez procesor kognitywny oparty na wiedzy w postaci reguł produkcji. EPIC stosowano do badania prawdopodobieństwa pomyłek i czasów reakcji, ale istnieje też wersja współpracująca z SOAR w zakresie planowania i rozwiązywania problemów, którą zastosowano do symulacji pracy kontrolerów lotu (Rosbe i inn. 2001).

SNePS (*Semantic Network Processing System*) używa do reprezentacji wiedzy w celu rozumowania i działania nie tylko podejścia logicznego, ale również ram i sieci semantycznych. System rozwijany jest od ponad 30 lat (Shapiro i inn. 2007) a jego celem jest zrozumienie natury procesów odpowiedzialnych za inteligencję przez eksperymentowanie z agentami poznawczymi, zdolnymi do rozwiązywania problemów i posługiwania się językiem naturalnym. Wiedza i przekonania agentów SNePS ma formę asercji dotyczących różnych obiektów, atrybutów i relacji. Pakiet wnioskowania SNIP (*SNePS Inference Package*), integruje rezultaty otrzymane za pomocą różnych form reprezentacji wiedzy, posługujących się odrębnymi mechanizmami wnioskowania (logicznym, opartym na ramach, oraz szukaniu dróg w sieciach semantycznych). Wykrycie sprzeczności przez system wymaga usunięcia lub zmiany przekonań na temat przyjętych założeń oraz propagacji tych zmian celem skorygowania błędnych wniosków uzyskanych w oparciu o takie założenia.



System kontroli SNePS Rational Engine planuje i podejmuje sekwencyjne działania korzystając z ram, a system analizy języka naturalnego oparty jest na klasycznej architekturze rozszerzonej gramatyki sieci przejść i analizatorze/syntezerze morfologicznym.

SNePS wykorzystano do analiz językowych i tworzenia odpowiedzi tekstowych, rozumowania zdroworozsądkowego, automatycznego rozszerzania słownika pojęć, sterowania symulowanymi agentami prowadzącymi konwersację z użytkownikami, w systemach pytań i odpowiedzi i innych zastosowaniach, a nawet teorii działania umysłu matematyka (uwzględniając enaktywne ugruntowanie symboli). Chociaż zademonstrowano interesujące wyniki wnioskowań system nie był dotychczas używany w rzeczywistych aplikacjach, trudno więc przewidzieć jaka będzie jego skalowalność dla dużych baz wiedzy. Program dostępny jest za darmo, napisany został w Common Lisp i ma użyteczny interfejs graficzny⁸.

NARS (*Non-Axiomatic Reasoning System*) to projekt rozwijany przez Pei Wang (2006) w ciągu ostatnich dwóch dekad⁹. Jest to ciekawa próba aproksymacji architektury poznawczej za pomocą niestandardowej logiki. System NARS przeznaczony jest do wnioskowania w oparciu o język reprezentacji wiedzy wykorzystujący semantykę ugruntowaną w pragmatyce (*experience-grounded semantics*). Logika nie-aksjomatyczna oznacza, że prawdziwość stwierdzeń logicznych oceniana jest na podstawie przeszłych doświadczeń systemu z podobnymi sytuacjami. Można ją wykorzystać do adaptacji w sytuacjach braku dokładnej wiedzy pozwalającej na jednoznaczne rozstrzygnięcia. Język reprezentacji, wraz ze zbiorem reguł wnioskowania, strukturami pamięci i mechanizmem kontrolnym pozwala traktować wykonywane zadania jako różne aspekty tego samego procesu. Wnioski mają również stopnie prawdziwości, oceniane na podstawie ich przydatności. Powstało kilka

⁸ <http://www.cse.buffalo.edu/sneps/>

⁹ <http://nars.wang.googlepages.com/>

prototypów systemów NARS o coraz większym stopniu wyrafinowania, jednak nie zademonstrowano jeszcze żadnego poważniejszego zastosowania.

ICARUS jest nowszą architekturą (Langley, 2005) przeznaczoną do sterowania agentami (robotami i awatarami). Zastosowano w niej reprezentację wiedzy w postaci reaktywnych umiejętności, każda z nich określająca jakąś sekwencję działań w kontekście określonych celów. Architektura obejmuje podsystem percepcyjny, planowania, wykonawczy i kilka podsystemów pamięci. Percepty i spostrzeżenia stoją w hierarchii nisko (odpowiadają początkowym fazom przetwarzania informacji zmysłowej), a pojęcia wysoko (odpowiadają rozpoznawanym obiektom, końcowej fazie). Mając dane percepty szuka się pasujących do nich pojęć, a mając dane cele szuka się reaktywnych umiejętności, które będą przydatne do ich spełnienia. Pamięć pojęć zawiera wiedzę o różnych klasach obiektów, związkach pomiędzy nimi, możliwościach manipulacji i stosowalności różnych procedur. Pamięć podzielona jest na długotrwałą (long-term memory, LTM) i krótkotrwałą (short-term memory, STM). Pamięć długotrwała zorganizowana jest hierarchicznie, pomagając w dopasowaniu perceptów, tworzeniu ich kombinacji pasujących do pojęć, jak i wyszukaniu odpowiednich umiejętności do wyboru działań w zależności od celu. Hierarchiczne, inkrementacyjne uczenie z krytykiem pozwala na propagowanie sygnałów nagrody wstecz w czasie, usprawniając osiągnięcie celów systemu na poziomie umiejętności i poszczególnych posunięć. W porównaniu z tradycyjnym uczeniem się z krytykiem zastosowane tu podejście hierarchiczne, dające priorytet akcjom o wysokim stopniu użyteczności, pozwala na znacznie szybsze uczenie się, a hierarchiczna organizacja obszernej pamięci pozwala na skupianie uwagi na zdarzeniach i obiektach w zasięgu sensorów, upraszczając procesy dopasowywania wzorców i redukując czasy reakcji systemu (Langley i Choi, 2006). ICARUS używa modułu planowania do konstrukcji drzewa zadań, ucząc się przy tym nowych koncepcji w sposób łatwy do zrozumienia, ale nowe koncepcje wysokiego poziomu nie jest łatwo tworzyć automatycznie.

Pokazano szereg interesujących zastosowań tej architektury do gier logicznych, poruszania się w symulowanym świecie, prowadzenia samochodu w symulowanych

warunkach miejskich jak i walk ulicznych agentów programowych. Brakuje w niej jeszcze równoległego przetwarzania pozwalającego na koordynację asynchronicznych sygnałów z sensorów, jak i radzenia sobie z informacją nieprecyzyjną.

3.2 Architektury emergentne

Ten rodzaj architektur kognitywnych związany jest w mniejszym lub większym stopniu z inspiracjami neurobiologicznymi. Propozycje tego typu rozwiązań pojawiły się już w książce McClellanda i Rumelharta (1986) w postaci modeli koneksjonistycznych. W modelach neuronowych elementy sieci reprezentują neurony i realizują proste funkcje, dopiero kwazistabilna konfiguracja pobudzeń tych elementów może być interpretowana jako ślad pamięci, np. rozpoznanie jakiegoś obiektu. Pomędzy takimi konfiguracjami istnieją skojarzenia, a więc możliwe są przejścia od jednej konfiguracji do drugiej. W modelach koneksjonistycznych mamy do czynienia z sieciami prostych elementów przetwarzających sygnały, a każdy z nich reprezentuje już jakieś pojęcie czy ślad pamięci, a więc pojedynczy element takiej sieci reprezentuje jakąś konfigurację pobudzeń wielu węzłów w sieciach neuronowych. Połączenia między węzłami sieci definiują możliwości przepływu aktywacji neuronowej od jednej konfiguracji do drugiej. W obu przypadkach elementy sieci oddziałują ze sobą zmieniając swój stan wewnętrzny i sposób oddziaływania z innymi elementami, a dynamika tych oddziaływań prowadzi do zmiany własności całej sieci i wyłaniania się (emergencji) nowych własności.

W modelach sieciowych mamy dwie możliwości organizacji pamięci: globalną, całkowicie rozproszoną, lub zlokalizowaną. Perceptrony Wielowarstwowe (Multi-Layer Perceptron, MLP) i inne sieci neuronowe używające nielokalnych funkcji transferu reprezentując informację w całkowicie rozproszony sposób: wszystkie parametry danej sieci mają wpływ na ostateczny wynik obliczeń. Generalizacja reakcji na nowe bodźce jest zwykle w takich sieciach całkiem dobra, ale próba nauczania nowej wiedzy może prowadzić do katastroficznej interferencji i zapominania poprzednio wyuczonych zależności (O'Reilly i Munakata, 2000). Sieci MLP inspirowane są przez przybliżony opis działania pojedynczych neuronów. Jeśli zwrócić uwagę na mikroobwody zamiast

pojedynczych neuronów to można przyjąć, że podstawową jednostką przetwarzającą informację nie są pojedyncze neurony, lecz fragmenty sieci, dokonujące projekcji sygnału na jakąś wybraną funkcję bazową, np. funkcję zlokalizowaną. Może to być odzwierciedleniem rezonansowych własności mikroobwodów, reagujących na specyficzne struktury w nadchodzących sygnałach. Rozwinięcia na funkcje bazowe, zwłaszcza zlokalizowane funkcje radialne (często są to funkcje Gaussa), są przykładem sieci przechowującej ślady pamięci w sposób zlokalizowany. Tylko nieliczne węzły sieci, które ulegają aktywacji przy danym pobudzeniu, mają wpływ na wynik jej działania. Sieci typu MLP można łatwo przekształcić w sieci modułarne, łącząc ze sobą pary neuronów i otrzymując w ten sposób filtry lokalizujące przepływ sygnału przez sieć, nie są to więc mocno odmienne podejścia (Duch i inn. 2001).

Metody uczenia architektur emergentnych są dość zróżnicowane (McClelland i Rumelhart 1986; O'Reilly i Munakata 2000). Ogólną zasadą jest kompresja informacji w wyniku uczenia.

Uczenie skojarzeniowe transformuje sygnały wejściowe w specyficzne, zwykle prostsze reprezentacje wyjściowe, które mogą służyć do kategoryzacji, heteroasocjacji, pamiętania pożądaných reakcji w określonym kontekście, lub też umożliwić dopełnianie brakującej informacji w niepełnych wzorcach wejściowych. Uczenie może być albo nadzorowane bezpośrednio przez oczekiwania, które po każdej prezentacji bodźców konfrontowane są z odpowiedziami sieci, lub też pośrednio, przez krytykę działania systemu po kilku krokach jego działania, jak to się dzieje w przypadku uczenia z krytykiem (reinforcement learning).

Uczenie konkurencyjne jest wynikiem wzajemnego hamowania się aktywnych jednostek, prowadząc do uproszczenia przepływających sygnałów, tworzenia się skupień podobnych pobudzeń w nienadzorowany sposób. Najprostsza forma takiego uczenia, znana jako „zwycięzca-bierze-wszystko” (winner-takes-all, WTA), pozostawia tylko jeden zwycięski element w stanie aktywnym, wyhamowując wszystkie pozostałe. Taki element służy za prototyp dla całej klasy sygnałów.

Uczenie korelacyjne, oparte na regule Hebba, pozwala zapisać korelacje statystyczne pomiędzy dochodzącymi sygnałami, tworząc uproszczony wewnętrzny model istotnych zależności w środowisku i pozwalając na wykrycie interesujących cech w sygnałach.

Architektury emergentne są bliższe problemom percepcji i wnioskowania na podstawie postrzeżeń, niż bardziej abstrakcyjnym formom rozumowania w oparciu o symbole. W zastosowaniu do reprezentacji złożonych form wiedzy i rozwiązywania problemów nie osiągnęły one jeszcze podobnego poziomu co architektury symboliczne, niemniej jest tu kilka propozycji wartych omówienia. Teoretycznie można by użyć każdego z dużych symulatorów neuronowych (np. Genesis, Neuron) do symulacji poszczególnych funkcji poznawczych, jednak budowa całej architektury kognitywnej wymaga dużych sieci, które muszą być oparte na znacznie prostszych modelach neuronów, niż modele oferowane przez duże symulatory.

IBCA (*Integrated Biologically-based Cognitive Architecture*) to architektura wzorowana na modelu procesów przetwarzania informacji przez kilka kluczowych regionów mózgu (O'Reilly i inn. 1999). Podkreśla się w niej rolę trzech głównych obszarów, realizujących różne typy pamięci: korę ciemieniową (posteriori cortex, PC), korę czołową (frontal cortex, FC), i formację hipokampa (HC). Organizacja sieci neuronowych w tych obszarach jest dość odmienna.

Kora ciemieniowa zawiera pola recepcyjne reagujące lokalnie, ale częściowo nakładające się, realizujące transformacje sensomotoryczne, skojarzenia, oraz hierarchiczne transformacje sygnałów o różnej modalności.

Kora czołowa używa izolowanych modułów z silną rekurencją, pozwalając na realizację pamięci roboczej, dzięki której można przerwać bieżące działanie, a potem do niego powrócić. Izolowane reprezentacje mogą się łączyć w kombinatoryczny sposób ze sobą, tworząc np. różnorodne kombinacje cech.

Hipokamp realizuje globalny model pamięci, w którym wszystkie wejścia pobudzają rzadkie reprezentacje (niewielka część neuronów jest aktywna), łączące różne cechy ze sobą, ale pozwalające na dyskryminację obiektów różniących się w nieznacznym sposób.

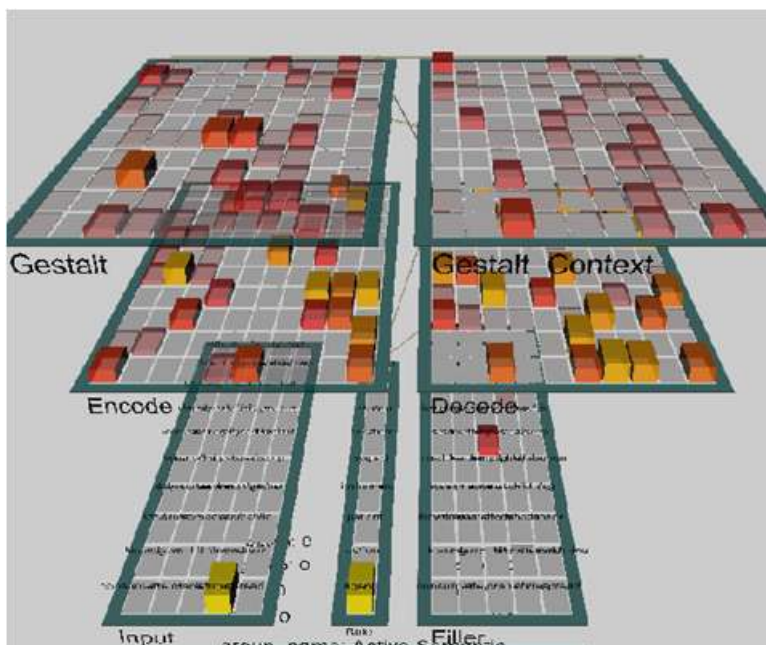
Moduł HC pozwala na realizację pamięci epizodycznej, zapisując aktywacje w PC i FC i pomagając uniknąć interferencji (zbytniego rozchodzenia się aktywacji) w PC.

Architektura IBCA uczona jest za pomocą algorytmu LEABRA (*Local, Error-driven and Associative, Biologically Realistic Algorithm*), czyli kombinacji Hebbowskiego uczenia korelacyjnego z konkurencyjnym hamowaniem w procesach typu „zwycięzca bierze większość”, oraz dwufazowej korekcji błędów, w sposób nieco lepiej biologicznie uzasadniony niż robi to algorytm wstecznej propagacji. Moduły PC i FC realizują powolne uczenie integrujące wiele ekspozycji na różne bodźce, pozwalając na odkrywanie regularności w środowisku i realizację zachowań senso-motorycznych. Moduł HC pozwala natomiast na szybkie uczenie, które zapamiętuje indywidualne epizody i pozwala na rozróżnianie poszczególnych zdarzeń bądź obiektów.

Współpraca pomiędzy uczeniem HC i FC/PC ilustruje działanie dwóch komplementarnych systemów uczenia przez mózgi. Pozwala to na rozwiązanie problemu konsolidacji wiedzy, transferu wiedzy z pamięci krótkotrwałej do długotrwałej, zapisującej „obraz świata” w uczonym modelu. IBCA działa współbieżnie, jest to model elastyczny, dający się przystosować do modelowania wielu form zachowań, wykazujący dobrą generalizację dzięki rozproszonym reprezentacjom. Wyższe czynności poznawcze (działania sekwencyjne, powiązania między elementami sytuacji lub obiektów o złożonej strukturze) pojawiają się w tym modelu dzięki pamięci roboczej implementowanej w module FC, która potrafi utrzymać przez jakiś czas i aktualizować aktywną reprezentację sytuacji. Dość skomplikowany, ale biologicznie dobrze umotywowany, mechanizm nagrody pozwala na antycypację nagrody przez rodzaj uczenia się z krytykiem.

Architektura IBCA zrealizowana została za pomocą symulatora Emergent (wcześniej PDP++), nadal aktywnie rozwijanego. Zrobiono za jej pomocą sporo modeli eksperymentów psychologicznych, takich jak test Stroopa, test sortowania kart Wisconsin, modelowano mechanizmy percepcji, różne rodzaje dysleksji i wiele innych zagadnień. Symulacje są mocno czasochłonne, gdyż modele neuronów, chociaż dość uproszczone w porównaniu z dokładniejszymi modelami biologicznymi, zawierają

bardzo wiele parametrów, np. trzy rodzaj kanałów jonowych. Procesy hamowania uwzględnia się w uproszczony sposób. Dla wielu symulacji funkcji poznawczych potrzeba licznych warstw oddziaływujących ze sobą neuronów, w rezultacie uczeniu podlega wiele tysięcy parametrów, powstaje więc problem skalowania całego systemu. W obecnej implementacji architektura dobierana jest przez użytkownika do problemu, a uczeniu podlegają tylko parametry sieci, ale nie jej struktura. Brakuje reprezentacji emocji, pozwalającej na wprowadzenie motywacji i ułatwiającej wybór celów, nie ma też koordynacji ruchu i wycucia czasu. Jest to bez wątpienia bardzo interesująca architektura pomagająca zrozumieć funkcje poznawcze, nie wiadomo jednak w jaki sposób można za jej pomocą osiągnąć podobny poziom w procesach rozwiązywania problemów i myślenia sekwencyjnego, jaki udaje się osiągnąć za pomocą architektur symbolicznych.



2 Przykład modelu rozumienia zdań w symulatorze Emergent.

NOMAD (*Neurally Organized Mobile Adaptive Device*) to architektura sterująca agentami, oparta na teorii „neuronalnego Darwinizmu” Geralda Edelmanna (1993, 1999). „Automaty Darwina”, czyli roboty sterowane przez tą architekturę, pokazują jak zasady

emergentne sprawdzają się w sterowaniu i rozpoznawaniu wzorców w czasie rzeczywistym. Informacja o świecie dostarczana jest przez kamery, mierniki odległości, sztuczne wibrysy (jak u szczura), zmysł dotyku i czujnik smaku (przewodności), oraz czucie proprioceptywne, przydatne do określenia położenia głowy i kierunku swojego ruchu. W symulatorze uwzględniono kilkanaście podobszarów mózgu, przetwarzających informację zmysłową i sterujących robotem. Jest on uczony w pseudo-naturalny sposób, tzn. rozwija się od „narodzin”, nic nie wiedząc, w sztucznym, ale zróżnicowanym środowisku.



3 Nomad w swoim środowisku rozpoznaje wzorce na klockach i kojarzy je z ich smakiem.

Eksperymenty z Nomadami pokazały rolę systemu wartości wykorzystującego wewnętrzne mechanizmy nagrody. System wartości robota to jego genetyczne predyspozycje do poszukiwania wrażeń i unikania sytuacji szkodliwych. Pozwala mu to na rozwinięcie interesujących form zachowań, podejmowania działań zwiększających

dopływ pozytywnych wrażeń, śledzenie obiektów, zbieranie „dobrze smakujących” klocków w jednym miejscu i unikanie klocków źle smakujących. Okazało się, że rozwój percepcyjny jest sprzężony z kontrolowaniem własnych ruchów. Formacja hipokampa jest kluczową strukturą w rozwoju pamięci epizodycznej jak i nawigacji przestrzennej (modele Darwin X-XI). Osiągnięto niezmienniczość wzrokowego rozpoznawania obiektów (Darwin VI-VII), łączenie cech złożonych obiektów w całość dzięki synchronizacji neuronów przez rekurencyjne połączenia w układzie wzrokowym¹⁰.

Symulacje tej architektury są bardzo wymagające, bo składa się ona z ponad 100 tysięcy neuronów i ponad 10 milionów synaps. Architektura Nomada jest w pełni współbieżna i działa w czasie rzeczywistym na wieloprocessorowych serwerach. Jest ona podstawą do budowy różnych „mózgo-podobnych urządzeń robotycznych” (Brain-Based Robotic Devices). Dotychczas nie pokazano jednak, jak ją rozszerzyć w stronę wyższych czynności poznawczych i symbolicznych procesów myślenia.

NuPIC (*Numenta Platform for Intelligent Computing*) jest dość nową propozycją architektury emergentnej opartej na Hierarchicznej Pamięci Temporalnej (Hierarchical Temporal Memory, HTM), która wynika z rozważań ogólnych nad algorytmem przetwarzania informacji przez mózgi (Hawkins i Blakeslee, 2004). Węzły sieci zorganizowane są w sposób hierarchiczny, motywowany przez rosnące rozmiary korowych pól recepcyjnych w obszarach zmierzających od pierwotnej kory zmysłowej, przez wtórną i wyższe obszary skojarzeniowe. Podobne efekty występują w architekturze IBCA, gdzie specyficzne połączenia pomiędzy warstwami prowadzą do coraz większych, niezmienniczych pól recepcyjnych, a w końcu do rozpoznania obiektu.

Węzły sieci HTM zorganizowane są w hierarchiczny sposób, każdym z nich ma pamięć i możliwości uczenia. W tym modelu podkreśla się temporalny aspekt percepcji, pamięć sekwencji wrażeń, która ułatwia rozpoznawanie i antycypację kolejnych wrażeń. Każdy poziom w hierarchicznej sieci uczony jest niezależnie by zapamiętać czasoprzestrzenne

¹⁰ Problem łączenia cech (the binding problem) wynika z tego, że różne cechy obrazu analizowane są przez wyspecjalizowane obszary w mózgu, nie ma jednego miejsca, w którym pojawia się reprezentacja obrazu, musi więc być jakiś mechanizm, który przypisuje rozproszone pobudzenia do jednego obiektu.

wzorce i rozpoznawać nowe, podobne wzorce dzięki współpracy procesów oddolnych i odgórnym (bottom-up/top-down), a więc bezpośredniej hierarchicznej analizie informacji i opartej na oczekiwaniach węzłów wyższego poziomu antycypacji ograniczającej interpretację bezpośredniej analizy. Architektura HTM ma wielu zwolenników, ale nie testowane jej jeszcze w aplikacjach na większą skalę.

Cortronics, lub architektura konfabulacji, to nowa architektura emergentna, zainspirowana przez pętlę wzgórzowo-korową w mózgu (Hecht-Nielsen, 2007). Pamięć zorganizowana jest w postaci modularnych sieci atraktorowych, nazywanych leksykonami. Każdy z nich składa się z kawałka kory nowej i połączonego z nią fragmentu kory wzgórza. Stabilne stany leksykonów nazywane są symbolami, każdy z nich reprezentowany jest przez grupę specyficznych neuronów. Nieliczne z nich są wspólne dla różnych par symboli, istnieje więc pewne nakrywanie pomiędzy leksykonami. Elementarna wiedza zapisywana jest w postaci połączeń pomiędzy neuronami różnych leksykonów. Wszystkie kawałki kory sumują się do całej kory, a kawałki kory wzgórza do części łączącej się z korą nową. Mechanizm konkurencyjnej aktywacji symboli zawartych w leksykonach, zwany konfabulacją, służy do antycypacji kolejnych stanów, ruchów czy słów. Jego działanie pozostawia tylko kilka neuronów w stanie aktywnym, tworząc symbol, który zwyciężył w konkurencji z innymi, lub też symbol zerowy, oznaczający „nie wiem”.

Sam proces konfabulacji nie wystarczy do realizacji procesów rozumowania czy reprezentacji złożonej wiedzy, dotychczas stosowany był jedynie do analizy tekstów. Konfabulacja jest interesującym procesem przydatnym nie tylko do antycypacji, lecz również realizacji wyobraźni i kreatywności (Duch & Pilichowski 2007; Duch 2007). Proces ten zachodzi w krótszej skali czasowej niż procesy rozumowania.

Inne architektury emergentne: idea globalnej przestrzeni roboczej, opisana przez Baarsa (1988), zrobiła dużą karierę, ale dotychczas zaimplementowano niewiele modeli na niej opartych. Shanahan (2006) opisał bardzo prostą implementację wykorzystującą bezwagowe sieci neuronowe (pRAM), stosując ją do sterowania symulowanym robotem. Inne implementacje tej idei dotyczą modeli hybrydowych i są opisane poniżej.

W ostatnich latach pojawiły się nowe propozycje architektur emergentnych, ale na razie niewiele wiadomo o ich własnościach z powodu braku dobrych symulatorów. Haikonen (2007) napisał książkę na temat świadomych maszyn, przedstawiając w bardzo ogólny sposób architekturę mózgu robota, brakuje w niej jednak konkretnych wyników. Anderson i jego koledzy sformułowali ciekawy projekt nazwany “Ersatz brain” (Anderson i inn. 2007), ale nie zbudowali jeszcze symulatora.

Idea autonomicznego rozwoju umysłu (autonomous mental development) ma już od dziesięciu lat silne wsparcie ze strony grupy robotyków (Weng i Hwang, 2006), zmierzając w podobnym kierunku co projekt Nomad Edelmana (1993; 1999) czy Cog Brooksa (1986; Brooks i Stein, 1994), czyli budowy działających w czasie rzeczywistym robotów, rozwijających się przez interakcję z otoczeniem. Korner i Matsumoto (2002), pracujący dla Hondy, argumentują, że emergentne architektury poznawcze powinny kontrolować ograniczenia, za pomocą których dokonuje się selekcji właściwego algorytmu z istniejącego repertuaru pozwalającego rozwiązać problem, a jeśli stereotypowe rozwiązanie nie jest możliwe, próbować stworzyć nowy algorytm. Ten pomysł zmierza w kierunku meta-uczenia, które jest jednym z najważniejszych zagadnień inteligencji obliczeniowej (Duch 2007a): rozwiązanie trudnego problemu wymaga nauczenia się, jakiej sekwencji transformacji trzeba dokonać by osiągnąć pożądaną cel.

DARPA zainicjowała program budowy biologicznie inspirowanych architektur kognitywnych (Biologically-Inspired Cognitive Architectures, BICA)¹¹. Wynikiem tego programu było szereg interesujących propozycji, np. rozszerzenia SOAR w kierunku inspirowanym przez neurobiologię, oraz projekt modelu ludzkiego umysłu (Comprehensive brain-based model of human mind; TOSCA, 2006), napisany wspólnie przez grupę ekspertów z wiodących instytucji amerykańskich. Można się więc będzie spodziewać znacznej aktywności w formułowaniu i implementacjach architektur emergentnych w bliskiej przyszłości.

¹¹ <http://www.darpa.mil/ipto/>

3.3 Architektury hybrydowe

Architektury symboliczne i emergentne w znacznej mierze się uzupełniają. Podejście symboliczne pozwala modelować wyższe funkcje poznawcze, takie jak planowanie i rozumowanie, w sposób przypominający pracę eksperta. Jednakże zamiana na symbole dużej ilości informacji, odbieranej przez zmysły wieloma kanałami w bardzo zróżnicowany sposób, w różnym kontekście, wymaga uwzględnienia procesów emergentnych, za pomocą których trudno jest zrealizować wyższe funkcje poznawcze. Dlatego systemy hybrydowe są najbardziej obiecującą drogą do budowy architektur kognitywnych, wykorzystując silne strony obu podejść do stworzenia architektur symulujących wszystkie etapy przetwarzania informacji przez mózgi (Sun i Alexandre 1997).

Architektury hybrydowe można z grubsza podzielić na dwa typy, lokalno-rozproszone, lub symboliczno-koneksjonistyczne, w zależności od rodzaju wykorzystywanych przez nie modułów pamięci. Do pierwszej klasy, bliższej systemom emergentnym, zaliczyć można architektury zawierające kombinację modułów pamięci lokalnej, w których każde pojęcie reprezentowane jest przez osobny węzeł, i pamięci rozproszonej, w której każde pojęcie reprezentowane jest przez zbiór częściowo nakrywających się węzłów. Do drugiej klasy zaliczymy kombinacje modułów symbolicznych, wykorzystujących reguły lub reprezentacje graficzne, z modułami koneksjonistycznymi, zarówno zlokalizowanymi jak i rozproszonymi (Sun & Alexandre 1997).

Mechanizmy uczenia również podzielić można na dwie kategorie (Sun i Zhang, 2004), w zależności od dominacji uczenia odgórnego (*top-down*) lub oddolnego (*bottom-up*). Uczenie odgórne oparte jest na przesyłaniu informacji od poziomu symbolicznego, na którym działają reguły logiczne lub inne mechanizmy wnioskowania, do subsymbolicznego, stanowiącego poziom dolny (pamięć rozproszona), który uczy się obserwując wyniki działań na wyższym poziomie. W uczeniu oddolnym wiedza zdobywana jest na poziomie subsymbolicznym i przekazywana do górnego poziomu w postaci reguł zawierających sformułowania koncepcyjnie na poziomie symbolicznym (Duch i inn. 2001, Sun i Merrill 2001).

Parę przykładów architektur hybrydowych omówiono poniżej, skupiając się na organizacji pamięci i mechanizmach uczenia.

ACT-R (*Adaptive Components of Thought-Rational*) jest rozwijaną od wielu dziesięcioleci architekturą kognitywną opartą na stworzonych przez Johna Andersona podstawach teoretycznych, dotyczących mechanizmów poznawczych (Anderson 1998; Anderson i Lebiere 2003)¹². Celem tego projektu jest budowa systemu, który będzie zdolny do wykonania wszystkich zadań poznawczych na zbliżonym poziomie i z podobnymi błędami co ludzie, oraz zrozumienie mechanizmów leżących u podstaw percepcji, myślenia i działania. Częścią centralną architektury ACT-R jest zbiór modułów do analizy percepcji, kontroli ruchu, moduły pamięci i dopasowania wzorców. Część percepcyjno-ruchowa dostarcza reprezentacji symbolicznych, służących jako interfejs pomiędzy częścią poznawczą a środowiskiem działania systemu. Dwa typy pamięci, deklaratywna i proceduralna, służą do przechowywania faktów i sposobów działania (procedur). Pamięć kodowana jest w symboliczno-koneksjonistycznych strukturach, za pomocą reguł produkcji na poziomie procedur, oraz porcji pamięci (chunks) na poziomie deklaratywnym, w postaci reprezentacji wektorowych zbioru własności, łącząc się z subsymbolicznym poziomem w postaci sieci koneksjonistycznej. Konstrukcje symboliczne (reguły produkcji i porcje pamięci) mają dodatkowe parametry oceniające ich przydatność i pozwalające na kontrolę sposobu ich użycia. Parametry te douczane są za pomocą probabilistycznych reguł Bayesa, co pozwala na określenie ich przydatności na podstawie przeszłych doświadczeń. Bufory pamięci ACT-R odpowiadają pamięci roboczej, w której dochodzi do dopasowania reguł do bieżącej sytuacji.

Zastosowano odgórną strategię uczenia: rezultatami złożonych operacji są konstrukty symboliczne, dzięki którym nie trzeba powtarzać szczegółowej drogi rozumowania w podobnej sytuacji w przyszłości (podobnie jak dzieje się to w SOAR). Nowy cel, aktywacja porcji pamięci deklaratywnej, lub identyfikacja postrzeganego obiektu, stają się obiektami w buforze pamięci, a system dopasowania wzorców ukierunkowany przez

¹² Strona projektu ACT-R <http://act-r.psy.cmu.edu>

procesy subsymboliczne poszukuje najlepszej reguły, która może pasować do znajdujących się w pamięci roboczej struktur. Mechanizm uczenia powoduje, że porcje pamięci, które są częściej aktywne, stają się łatwiej dostępne i są preferowane. Również reguły produkcji, które okazały się bardziej przydatne, mają większą oczekiwaną użyteczność i są częściej wybierane w sytuacjach, w których można zastosować wiele reguł.

Architekturę ACT-R można z grubsza powiązać z różnymi funkcjonalnie określonymi obszarami mózgu, chociaż jest to architektura wywodząca się bardziej z inspiracji psychologicznych niż neurobiologicznych. W ostatnich latach podjęto próby powiązania aktywności mózgu mierzonej za pomocą fMRI z aktywacją modułów ACT-R w złożonych zdaniach poznawczych. Liczna grupa jej użytkowników zastosowała modele zrobione w ACT-R do zagadnień dotyczących symulacji bardzo wielu eksperymentów w psychologii poznawczej. ACT-R jest też podstawą budowy inteligentnych systemów wspomagających nauczanie, produkowanych przez komercyjną firmę Carnegie Learning¹³, której produkty używane są przez pół miliona uczniów w USA.

CLARION (*The Connectionist Learning Adaptive Rule Induction ON-line*) jest rozwijaną od ponad 10 lat architekturą hybrydową, mającą na celu rozwój agentów kognitywnych przydatnych do różnych zadań, jak i zrozumienie procesów zachodzących w mózgu w czasie uczenia i rozwiązywania problemów (Sun i Alexandre, 1997; Sun i Zhang, 2004). Składają się na nią cztery główne podsystemy: podsystem motywacji (MS), podsystem działania (action-centered subsystem, ACS), podsystem niezwiązany z działaniem (non-action-centered subsystem, NCS), i podsystem meta-poznania (metacognitive subsystem, MCS). W każdym z nich odróżnić można dwojaki reprezentacje, jawne (symboliczne) i utajone (subsymboliczne). System motywacji dostarcza celów dla analizy percepcji i kontroli działań, ACS kontroluje działania agenta, NCS odpowiada za wiedzę ogólną systemu, a MCS kontroluje całość. Pamięć zlokalizowana odpowiada za wiedzę symboliczną, a pamięć rozproszona za wiedzę utajoną (implicit knowledge). Do każdego rodzaju wiedzy stosowana jest inna strategia

¹³ <http://carnegielearning.com>

uczenia. Pamięć rozproszona korzysta zarówno z uczenia nadzorowanego jak i uczenia z krytykiem – zastosowano tu algorytm Q-uczenia zaimplementowany dla sieci MLP (Sun i inn. 2001). Wiedza zdobyta na tym poziomie jest zamieniana na wiedzę symboliczną na poziomie jawnym. Stosowane jest też uczenie odgórne, w którym moduły pamięci rozproszonej obserwują rezultaty działania reguł na wysokim poziomie, co pozwala wstępnie utworzyć prawidłowe zachowania i zainicjalizować parametry sieci (Sun i Zhang, 2004). CLARION zaczyna więc od pewnej wiedzy ogólnej, ale ucząc się będzie coraz bardziej polegać na wiedzy zdobytej na niższym, utajonym poziomie.

Architektura stosowana była do symulacji wielu eksperymentów poznawczych, ale również w ciekawej aplikacji dotyczącej nawigacji pojazdu na zaminowanym polu, demonstrując swoje możliwości podejmowania złożonych decyzji.

DUAL (Nester i Kokinov, 2004) to dość nowa architektura oparta na ideach “Społeczeństwa umysłu” Marvinina Minsky’ego (1986). Jest to wieloagentowa architektura hybrydowa, z jednolitymi reprezentacjami mentalnymi, strukturami pamięci i mechanizmami przetwarzania informacji kontrolowanymi przez współdziałające ze sobą mikroagenty. DUAL nie ma centralnego kontrolera, w rezultacie przepływ informacji ciągle się w nim zmienia, dostosowując do sytuacji i wspomagając emergencję nowych funkcji. Współdziałający agenci tworzą większe kompleksy, koalicje i formacje, a bardziej przydatne z nich mogą ulec reifikacji. Działanie systemu można rozpatrywać na różnym poziomie szczegółowości, na najniższym poziomie mikroagentów, na mezopoziomie emergentnych, dynamicznych koalicji tych mikroagentów, lub na makropoziomie działania całego systemu, na którym możliwa jest interpretacja psychologiczna jego własności. Mikro-ramy używane są do reprezentacji faktów, a ważkość tych faktów w danym kontekście reprezentowana jest w połączeniach i oddziaływaniach pomiędzy elementami sieci, w której rozchodząca się aktywacja zmienia dostępność informacji zawartej w jej węzłach. Powiązania pomiędzy mikroagentami wynikają z zawartości szufladek ich ramek, a wagi połączeń kontrolują wzajemny wpływ agentów na siebie.

Architektura DUAL została wykorzystana w modelach rozumowania przez analogię (model AMBR), rozumowania dedukcyjnego, analizie oddziaływania pomiędzy percepcją, pamięcią i rozumowaniem analogicznym, w modelu pamięci epizodycznej, ocen dokonywanych przez ludzi, modelowania roli kontekstu i efektów torowania na dynamikę procesów poznawczych. Bez wątpienia jest to przydatna architektura, która może wyjaśnić wiele procesów poznawczych. Nie jest jednak jasne, na ile da się ją przeskalować by rozwiązać problemy dotyczące złożonego rozumowania na poziomie eksperta.

LIDA (*The Learning Intelligent Distribution Agent*) jest nowszym projektem, opartym na teoretycznych podstawach budowy “świadomych” agentów programowych, wykorzystujący teorię globalnej przestrzeni roboczej Baarsa (Franklin, 2006). Jest to nowsza wersja architektury IDA, którą zastosowano do automatyzacji procesu przypisywania marynarzy o odpowiednich kwalifikacjach do pracy na statkach Marynarki Wojennej (US Navy).

LIDA stosuje mieszaną, symboliczno-koneksjonistyczną organizację pamięci, próbując ugruntować sens wszystkich symboli na poziomie subsymbolicznym poprzez działania w środowisku, jak to opisali Brooks i Stein (1994). LIDA ma odrębne moduły dla percepcji, wyboru działania, pamięci roboczej i pamięci semantycznej, antycypacji i uczenia się zadań proceduralnych, spełniania ograniczeń, rozważań i negocjacji, rozwiązywania problemów, emocji, meta-poznania i zachowania podobnego do świadomego (conscious-like behavior). Większość podstawowych, prostych operacji wykonywana jest przez codelety, wyspecjalizowane fragmenty sieci spełniające rolę nieświadomych procesorów przetwarzających informację w modelu globalnej pamięci roboczej (Baars, 1988). Uczenie percepcyjne, epizodyczne i proceduralne sterowane jest oddolnie. Uczenie percepcyjne dotyczy rozpoznawania nowych obiektów, kategorii, relacji, i oparte jest albo na zmianie siły połączeń synaptycznych między węzłami, albo na tworzeniu nowych węzłów i połączeń w pamięci percepcyjnej. Uczenie epizodyczne oznacza zapamiętywanie specyficznych zdarzeń: co, gdzie, kiedy, zdarzeń pojawiających się w pamięci roboczej, a więc dostępnych świadomości. Uczenie

proceduralne, czyli uczenie się nowych działań i sekwencji działań potrzebnych do rozwiązania postawionych problemów, realizowane jest na dwa sposoby. Pierwszym jest selekcja działań ze znanego repertuaru, a drugim konstrukcja nowych reprezentacji dla sekwencji działań za pomocą uczenia z krytykiem.

Jest to bardzo ciekawa propozycja znajdująca się nadal w stadium intensywnego rozwoju. Na tym etapie trudno jest jeszcze przewidzieć, jaki poziom kompetencji uda się tą drogą osiągnąć w zagadnieniach związanych z językiem, widzeniem czy rozumowaniem opartym na perceptach.

Polyscheme (Cassimatis 2007) jest ciekawą architekturą, która próbuje zintegrować wiele metod reprezentacji wiedzy, rozumowania i schematów wnioskowania, przydatnych do rozwiązywania problemów. W tej architekturze używa się pojęcia „specjalistów” (nie do końca odpowiadającego codeletom czy mikroagentom), których zadaniem jest modelować różne aspekty świata używając specyficznych reprezentacji i metod wnioskowania, oddziałując z innymi specjalistami i ucząc się od nich. Wiedza reprezentowana jest za pomocą sieci neuronowych, grafów ograniczeń, asercji logicznych, ram i skryptów. Zadaniem specjalisty zajmującego się uwagą jest doprowadzić do skupienia się nad wybranymi schematami wnioskowania w danym kontekście, poprzez dopasowanie wzorców, skryptów, procesów szukania, stochastycznych symulacji lub rozumowania kontrfaktycznego. Operacje używane w algorytmach rozwiązywania problemów, takie jak tworzenie podcelów, wnioskowanie w przód, tworzenie alternatywnych reprezentacji problemu czy transformacje sprawdzające identyczność różnych reprezentacji, wykonywane są przez specjalistów dysponujących różnymi formami reprezentacji odnoszącymi się do tych samych aspektów.

Polyscheme można używać zarówno do rozumowania abstrakcyjnego, w niektórych zadaniach możliwa też integracja procesów percepcji i kontroli działania niższego poziomu, dzięki czemu ta architektura nadaje się do sterowania fizycznymi agentami (robotami). Zastosowania obejmują modelowanie eksperymentów psychologicznych z rozumowaniem niemowląt na temat tożsamości obiektów, przyczynowości, relacji

przestrzennych. Polyscheme jest architekturą wykorzystującą meta-uczenie. Jej dalszy rozwój będzie ważnym krokiem w kierunku zrozumienia mechanizmów poznawczych.

4CAPS (Just i Varma 2007) jest unikalną architekturą, zaprojektowaną do wykonywania złożonych zadań, takich jak rozumienie języka naturalnego, rozwiązywanie problemów i rozumowanie przestrzenne. Jej unikalną własnością jest możliwość porównania aktywności modułów 4CAPS z wynikami obserwacji aktywności mózgu, mierzonymi za pomocą fMRI. U podstaw projektu leży przekonanie, że myślenie jest wynikiem jednoczesnej aktywności wielu obszarów mózgu współpracujących ze sobą, które można modelować za pomocą modularnych sieci neuronowych. Prowadzi to do architektury w której podsystemy (moduły) odpowiadają poszczególnym obszarom mózgu przetwarzającym informację w różny sposób. Np. obszar Wernickiego specjalizuje się w konstruowaniu i selektywnym dostępie do ustrukturalizowanych, hierarchicznych reprezentacji sekwencyjnych. Każdy moduł może być częścią większego systemu wykonującego wielorakie funkcje poznawcze, ale ma ograniczone zdolności obliczeniowe wpływające na jego dostępność do wykonywania zadań. Funkcje przypisywane są różnym modułom w zależności od dostępności zasobów, a więc topologia całej sieci rozwiązującej jakiś problem nie jest ustalona, to samo zadanie można rozwiązać na różne sposoby.

Program napisany jest w Lispie i można prowadzić z nim własne eksperymenty. W obecnej postaci architektura 4CAPS nie jest nastawiona na osiągnięcie wysokiego poziomu inteligencji a jedynie na odtworzenie wybranych rezultatów badań behawioralnych. Należy do nich badanie szybkości odpowiedzi i liczby błędów przy rozwiązywaniu problemów, rozumowaniu przez analogię, współpracy człowiek-maszyna, rozwiązywaniu problemów, rozumieniu dialogu i w innych złożonych zadaniach, rozwiązywanych zarówno przez ludzi zdrowych jak i niepełnosprawnych umysłowo (Just i Varma, 2007).

The Novamente AI Engine (Goertzel, 2006) jest oparta na ideach dotyczących działania złożonych systemów, dynamiki procesów umysłowych i emergentnych wzorców, wywodzących się z filozofii umysłu podkreślającej rolę wzorców i

emergentnych struktur (patternist philosophy of mind), pierwotnie naszkicowanej w modelu *psynet* (Goertzel 1997). Podobnie jak w „społeczeństwie umysłów” Minskiego i globalnej przestrzeni roboczej Baarsa, stany mentalne pojawiają się w wyniku samoorganizacji i oddziaływań związanych z realizacją celów, które wpływają na wzorce pobudzeń tworzące stany mentalne. Emergentne własności aktywacji sieci powinny doprowadzić do hierarchicznej i relacyjnej (hieterarchicznej) organizacji wzorców mentalnych. Wnioskowanie oparte jest na wersji logiki probabilistycznej (probabilistic term logic, PTL) i Bayesowskim algorytmie optymalizacji (BOA). Działania, percepty i stany wewnętrzne reprezentowane są przez struktury drzewiaste.

Jest to nadal architektura eksperymentalna, ale szybko rozwijana przez firmę komercyjną. Pojawiły się pierwsze próby jej zastosowania do sterowania wirtualnymi zwierzakami, zdolnymi do uczenia się na podstawie w miarę naturalnych interakcji w wirtualnym świecie (Goertzel, 2008).

Shruti (Shastri i Ajjanagadde 1993) jest biologicznie motywowaną architekturą kognitywną przeznaczoną głównie do modelowania rozumowania refleksyjnego. Koneksjonistyczna sieć reprezentuje tu obiekty, relacje, typy i reguły przyczynowe. Można w niej wydzielić klastry kodujące uniwersalne lub egzystencjalne kwantyfikatory, jak też uwzględnić stopnie przekonania o prawdziwości zakodowanych faktów. Węzły wysyłają impulsy o różnej częstotliwości, które mogą zsynchronizować pracę odległych węzłów dynamicznie wiążąc je ze sobą, co pozwala na reprezentację złożonej wiedzy i skomplikowane wnioski.

Jest to architektura o wielkim potencjale, ale pomimo stosunkowo długiego czasu rozwoju nie znalazła jeszcze żadnych poważnych zastosowań do analizy tekstów lub rozwiązywania problemów.

4. Perspektywy

Jak widać z powyższego przeglądu jest już wiele interesujących architektur kognitywnych o dużych możliwościach rozwoju, stwarzających nadzieję na budowę sztucznego umysłu. Niektóre z wymienionych powyżej projektów rozwijały się przez

dziesięciolecia, inne są stosunkowo nowe. Powoli powstaje nowa dziedzina, komputerowa psychologia poznawcza, pozwalająca na znacznie głębsze zrozumienie procesów poznawczych niż było to możliwe w psychologii opisowej i eksperymentalnej. Mimo znacznej liczby publikacji na temat architektur kognitywnych niewiele jest jeszcze rzeczywistych zastosowań na szerszą skalę (wyjątkiem jest tu ACT-R). Ważnym kierunkiem jest więc próba przeskalowania demonstracyjnych aplikacji na bardziej ambitne zastosowania, tworzenie komputerowych partnerów, zdolnych do wspierania i codziennej współpracy z ludźmi.

4.1 Jak to oceniać?

Warto definiować nowe wyzwania, ambitne, ale nie beznadziejnie trudne problemy, na których można testować ograniczenia architektur kognitywnych. Integracyjnymi modelami działania człowieka interesują się różne gałęzie przemysłu maszynowego. Porównanie systemów agentowych modelujących zachowanie pozwoliło zebrać wyniki obrazujące decyzje ludzi i programów w uproszczonym środowisku kontroli lotów (Gluck i Pew, 2005). Sporo wysiłku włożono w rozwój metod porównywania agentów programowych. Na warsztatach AAAI "Evaluating Architectures for Intelligence" (Kaminka i Burghart, 2007) dyskutowano na temat środowisk do testowania zachowań autonomicznych kierowców w warunkach miejskich, a także sposobów oceny elastyczności i adaptacyjności składowych agentów mających wykazywać się ogólną inteligencją.

Jedną z możliwych ocen architektur kognitywnych mających ambicję do osiągnięcia ludzkich kompetencji jest mierzenie „kognitywnego wieku”. W tym celu trzeba ustalić, jakiego rodzaju zachowania, typowe dla ustalonego wieku dzieci, dany system powinien symulować. Problemy można podzielić na wiele precyzyjnie określonych grup, dotyczących percepcji słuchowej i wzrokowej, rozumienia języka, rozumowania zdroworozsądkowego, rozumowania abstrakcyjnego, ogólnej wiedzy o świecie, uczenia się, rozwiązywania problemów, wyobraźni, kreatywności. Rozwiązanie wszystkich problemów z danej grupy wiekowej upoważniłoby wówczas dany system do stwierdzenia, że osiągnięty został poziom inteligencji odpowiadający temu wiekowi.

Można się spodziewać, że w pewnych obszarach postęp będzie szybszy niż w innych, np. programom łatwiej jest osiągnąć dobre wyniki dla rozumowania abstrakcyjnego niż dla percepcji wzrokowej. Z tego punktu widzenia w wybranych obszarach niektóre architektury okażą się dorosłe, a w innych będą na poziomie niemowląt. Eksperci zajmujący się inteligencją zgadzają się zwykle z propozycją Gardnera (1993) by wyróżnić siedem (lub więcej) rodzajów inteligencji: logiczno-matematyczną, językową, orientacji przestrzennej, muzyczną, kinestetyczną, interpersonalną i intrapersonalną. Można do tego dodać inteligencję emocjonalną.

Trudno jest zgromadzić ogólną wiedzę o świecie, ale jej testowanie można zrobić porównując odpowiedzi programów z odpowiedziami dzieci w określonym wieku. Jeśli dany system odpowie na wszystkie pytania, na które zna odpowiedź 5-latek, to można mu przypisać taki wiek w zakresie ogólnej wiedzy o świecie. Bazy wiedzy są zwykle ograniczone do specjalistycznych dziedzin i nie ma w nich informacji, na które potrafi odpowiedzieć dziecko. Baza systemu Cyc jest tu wyjątkiem (Panton i inn. 2006), a ponieważ jest w znacznej części publicznie dostępna można próbować użyć zawartej w niej wiedzy w innych systemach.

Taka analiza z pewnością powinna pomóc zrozumieć, jakiego rodzaju inteligencji spodziewać się można po projektach robotycznych skoncentrowanych na podejściu rozwojowym, podkreślającym rolę ciała, a jakiego rodzaju ograniczenia mają podejścia symboliczne, które osiągną wyższy wiek mentalny w całkiem innych kategoriach. Brooks słusznie zauważył, że słonie nie grają w szachy, a uważamy je za inteligentne zwierzęta (Brooks 1986), a później wyraził nadzieję (Brooks i Stein 1994), że roboty dysponujące zintegrowaną percepcją, wzrokiem, słuchem, zdolnościami do manipulacji, kontrolowane przez masowo równoległe wieloprocessorowe komputery „nauczą się myśleć wykorzystując swoje doświadczenia działania by stopniowo rozwiązywać coraz bardziej abstrakcyjne zadania”. Projekt Cog, rozwijany przez jego grupę od prawie dwóch dziesięcioleci, dorobił się licznych naśladowców, ale pozostał na poziomie agenta reaktywnego i nie widać w jaki sposób miałyby się w tak zbudowanym robocie rozwinąć abstrakcyjne myślenie. Słonie nie grają w szachy, ale na pewno nie nauczą się

tęgo precyzyjne wywijając trąbą. Behawioralna inteligencja w robotyce może wymagać ucieleśnienia, ale inteligencja abstrakcyjna, przydatna matematykowi, niekoniecznie.

Jest całkiem możliwe, że zagadnienia dotyczące percepcji najlepiej rozwiążą systemy neuromorficzne, ściśle wzorowane na strukturze układu wzrokowego. Osiągnięcie wysokiego stopnia inteligencji w zadaniach dotyczących percepcji może wymagać dość szczegółowych modeli. Jednak by się tego dowiedzieć potrzebujemy specyficznych testów.

4.2 Tendencje rozwojowe i nowe pomysły

Przegląd architektur kognitywnych pokazuje liczne trendy, które prawdopodobnie zdominują rozwój tej dziedziny w najbliższej przyszłości. Biologiczne inspiracje (BICA) stają się coraz ważniejsze i nawet zwolennicy tradycyjnych, symbolicznych systemów takich jak SOAR, planują liczne rozszerzenia inspirowane przez badania nad mózgiem. Na razie pomijane są regulacyjne możliwości pnia mózgu, które mogą dostarczyć meta-kontroli przez wybór odpowiedniego typu zachowania i możliwości kontroli uwagi, za to jest duże zainteresowanie analizą emocji, syntezą zachowań emocjonalnych przez awatary i roboty oraz wykorzystaniem emocji w procesach decyzyjnych.

Można zaproponować wiele architektur typu BICA, wydaje się jednak, że w przyszłych propozycjach zachowane zostaną pewne kluczowe cechy. Jedną z nich jest użycie wielu różnych podsystemów pamięci, jak to się już dzieje w różnych architekturach kognitywnych wszystkich typów. Przetwarzanie mowy lub tekstu wymaga rozpoznania podstawowych elementów, mapowania dźwięków na litery, fonemy, ciągi liter, słowa, pojęcia, rozstrzygnięcie dwuznaczności przez ustalanie znaczenia pojęć w jakiejś ontologii, wreszcie semantycznej reprezentacji całości, zrozumienia w kontekście całej wiedzy o świecie. Te kroki wymagają odniesienia do różnego rodzaju pamięci o dość odmiennej strukturze i własnościach.

Pamięć rozpoznawcza pozwala zwrócić uwagę na niespodzianki, nieregularności, dziwną pisownię lub brzmienie słowa, które może być pomyłką ale może być też czymś

nowym, specyficzną nazwą, obcym słowem, próbą uniknięcia filtrów antyspamowych w systemach poczty elektronicznej. Potrafimy rozpoznać czy już coś widzieliśmy dla bardzo wielkiej liczby struktur, o wiele większej niż potrafimy zapamiętać. Pamięć rozpoznawczą można zaimplementować w prostych sieciach bez ukrytych warstw lub w macierzach korelacji (Duch i Pilichowski 2007). Istniejące architektury nie wymieniają pamięci rozpoznawczej jako jednego z istotnych elementów.

Pamięć semantyczna służy nie tylko jako hierarchiczna ontologia, pozwalająca porządkować pojęcia, ale w mózgu jest podstawą procesów rozchodzenia się aktywacji, uaktywniania kontekstu, dostarczając wiedzy, która jest konieczna do ujednoznacznienia pojęć. Stosowane dotychczas ontologie tworzone są przez specjalistów, używają mało znanych pojęć (np. biologicznych taksonomii zwierząt czy roślin), nie opisują więc wiedzy potocznej. Modele pamięci semantycznej nie zawierają też wiedzy strukturalnej (krzesła mają różne formy, które możemy sobie wyobrazić, ale ich struktura nie jest opisana w ontologii). Brakuje w nich też wiedzy relacyjnej, która wynika z kontekstowego uczenia się (krzesło – stół – siadanie – jadalnia ...). Ontologia Cyc oparta na ramach zawiera bardzo dużo informacji, ale trudno jest z niej korzystać w sprawny sposób.

Inspiracje neurobiologiczne w zagadnieniach związanych z percepcją pojawiają się w metodach analizy sygnału przez układ słuchowy, wzrokowy czy węchowy. Znacznie gorzej wygląda kwestia neurobiologicznych inspiracji w przypadku wyższych czynności poznawczych. Lamb (1999) sformułował co prawda neurolingwistyczny manifest „Jest to próba zrozumienia procesów odpowiedzialnych za rozumienia języka w ludzkim mózgu, systemu, który umożliwia nam mowę i pismo, rozumienie, myślenie ...”.

Proponowane przez niego podejście ogranicza się na razie do rozważań różnych ciekawostek lingwistycznych i nie zostało nigdy użyte do analizy tekstu. Ogólna idea śledzenia rozchodzenia się aktywacji w sieciach neuronowych kodujących pojęcia językowe wymaga uzupełnienia o praktyczny algorytm modelowania „ścieżek rozchodzenia się aktywacji w mózgu”. Taki algorytm można oprzeć na ontologiach zawierających liczne relacje pomiędzy pojęciami, nie wystarcza tu ontologie

hierarchiczne. Jedynie ontologie medyczne zawierają wiele relacji, dzięki czemu da się prześledzić, jak powinny się rozchodzić w mózgu eksperta aktywacje neuronów, tworząc wektorowe lub sieciowe reprezentacje pojęć (Duch i inn. 2008). Otwiera to drogę do konstrukcji sieci semantycznych i reprezentacji wektorowych na wielką skalę, pozwalając na aproksymację wiedzy symbolicznej i specyficznych skojarzeń z nią związanych. Próba zbudowania opisu pojęć na podstawie analizy ontologii, słowników elektronicznych, encyklopedii, aktywnego szukania informacji i rezultatów projektów kolaboracyjnych opisana została w (Szymański i inn. 2008). W tej pracy pamięć semantyczna wykorzystana została do sterowania awatarem w grach słownych.

Pamięć epizodyczna konieczna jest po to, by zapamiętać bieżący kontekst, interpretować zdarzenia odwołując się do historii i obecnej sytuacji. Różne elementy pamięci semantycznej i epizodycznej przydatne w danym momencie znajdują się w **pamięci roboczej**. Wszystkie rodzaje pamięci ściśle ze sobą współpracują. Rozpoznawanie jest łatwiejsze, jeśli się wie, czego się szuka, a to wynika z bieżącego epizodu i pojęć w pamięci semantycznej, których aktywność wzrosła dzięki efektorowania, wynikłych z wcześniejszego dialogu czy kontekstu, antycypacji mogących się pojawić pojęć i hamowania aktywności pojęć niepasujących do kontekstu. Pamięć epizodyczna tworzy się przy wykorzystaniu relacji semantycznych.

Chociaż istnieją propozycje architektur kognitywnych oparte głównie na pamięci (Hoya, 2005), to rola różnych typów pamięci nie została w nich uwypuklona. Jedynie emergentna architektura IBCA (O'Reilly i inn, 1999) wiąże trzy ważne obszary mózgu z różnymi funkcjami pamięci, pomijając jednak pamięć rozpoznawczą.

Symboliczne podejście do języka naturalnego jest marnym substytutem dla procesów neuronowych, odpowiedzialnych za rozumienie mowy. Nieco lepsze przybliżenie oferuje model wektorowy, w którym zapisać można chwilową aktywność różnych grup neuronów kodujących powiązane pojęcia. W praktyce model wektorowy nie formułuje się jednak jako przybliżenie do zachodzących w mózgu procesów a jedynie jako statystycznie uzasadnioną próbę uwzględnienia kontekstu pojęć, określając korelacje pomiędzy pojawianiem się różnych pojęć w dużych korpusach tekstowych (Manning i

Schütze, 1999). Taka aproksymacja nie oddaje jednak wielu istotnych cech procesów percepcji-nazywania-działania zachodzących w mózgu (Dehaene i inn. 2005; Meeter i Murre, 2005). Rozważania na temat praktycznych algorytmów tworzenia i wykorzystywania reprezentacji pojęć są w (Duch i inn, 2008; Duch 2008).

W kilku architekturach zastosowano zbiór współpracujących ze sobą mikrofunkcji (mikroagentów, codeletów, specjalistów), które oddziaływując ze sobą wytwarzają stany emergentne, traktowane jako elementy wyższego rzędu. Spojrzenie na działanie mózgu z tego punktu widzenia, definiując na różnym poziomie elementy bazowe i ich oddziaływania, od pojedynczych neuronów do współpracujących ze sobą mózgów na poziomie interakcji społecznych, przedstawione zostało w pracy (Duch i Mandziuk, 2004). Pojedyncze neurony mają bardzo ograniczoną wiedzę, modelowe neurony progowe zwykle jeden parametr (próg pobudliwości), a ich interakcje przebiegają ze zmienną siłą dzięki połączeniom wewnątrz sieci. Grupy neuronów ściśle ze sobą współpracujących tworzą wyspecjalizowane procesory, posiadające znacznie bogatszą wiedzę wewnętrzną (stany wewnętrzne), oddziałując z innymi procesorami przekazując im złożoną informację. Takie procesy można przybliżyć za pomocą systemów wieloagentowych, które poprzez oddziaływania ze sobą mogą tworzyć koalicje zdolne do specyficznego przetwarzania informacji, reprezentujące większe obszary kory i ośrodki podkorowe mózgu. Funkcje, które spełniają te obszary, mogą się zmieniać w dynamiczny sposób, formując nowe metody rozwiązywania problemów w razie potrzeb. Takie idee meta-uczenia pozwalają na przewyciężenie problemów o dużej złożoności kombinatorycznej, pozwalając rozwiązywać problemy, dla których znane są tylko częściowe obserwacje i trzeba utworzyć całościowy model (Duch 2007). Prace nad ogólnym systemem do dogłębnej analizy danych, wykorzystującym idee meta-uczenia, są w toku (Grąbczewski i Jankowski, 2007).

Rola wyobraźni, kreatywności i używania wiedzy w intuicyjny sposób, oraz rola nie-dominującej półkuli mózgu w rozwiązywaniu problemów wymagających wglądu oraz interpretacji pojęć opisana została w pracach (Duch i Pilichowski 2007; Duch 2007). Planowana jest integracja tych pomysłów w formie architektury kognitywnej (Duch

2005), która będzie specjalizować się przede wszystkim w zadaniach wymagających rozumienia języka naturalnego.

Literatura

- Anderson J.A, P. Allopenna, G.S. Guralnik, D. Sheinberg, J.A. Santini, Jr., D. Dimitriadis, B.B. Machta, and B.T. Merrit, Programming a Parallel Computer: The Ersatz Brain Project. W: Duch W, Mandzuik J. (red.), Challenges to Computational Intelligence. Springer: Berlin, pp. 61-88, 2007.
- Anderson J.R. (1998), *Uczenie się i pamięć. Integracja zagadnień*. Warszawa: WSiP
- Anderson J.R, Lebiere C. (2003), The Newell test for a theory of cognition. *Behavioral and Brain Science* **26**, 587-637.
- Anderson, J.R. (2007), *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.
- Baars B.J. (1988), *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
- Brooks R. (1986), Elephants don't play chess. *Robotics and Autonomous Systems* **6**, 3-15.
- Brooks R., Stein L.A. (1994), Building Brains for Bodies. *Autonomous Robotics* **1**, 7-25.
- Carpenter R, Freeman J. (2005), Computing Machinery and the Individual: the Personal Turing Test, paper available at <http://www.jabberwacky.com/>.
- Cassimatis N.L. (2007), Adaptive Algorithmic Hybrids for Human-Level Artificial Intelligence. *Advances in Artificial General Intelligence*. IOS Press. Eds. B. Goertzel, P. Wang, s. 94-110.
- Dehaene S, Cohen L, Sigman M, Vinckier F. (2005), The neural code for written words: a proposal. *Trends in Cognitive Science* **9**, 335-341.

- Duch W (1994) A solution to the fundamental problems of cognitive sciences. International Philosophical Preprint Exchange, WWW Archive.
- Duch W (1996) From cognitive models to neurofuzzy systems - the mind space approach. *Systems Analysis-Modelling-Simulation* 24 (1996) 53-65
- Duch W. (2005), Brain-inspired conscious computing architecture. *Journal of Mind and Behavior* 26(1-2), 1-22.
- Duch W. (2007), Intuition, Insight, Imagination and Creativity. *IEEE Computational Intelligence Magazine* 2(3), 40-52.
- Duch W. (2007a), Towards comprehensive foundations of computational intelligence. In: W. Duch and J. Mandziuk, *Challenges for Computational Intelligence*. Springer Studies in Computational Intelligence, **63**, s. 261-316.
- Duch W, Adameczak R., Grąbczewski K. (2001), A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Transactions on Neural Networks*, **12**, 277-306.
- Duch W, Mandziuk J. (2004), Quo Vadis Computational Intelligence? In: *Machine Intelligence. Quo Vadis? Advances in Fuzzy Systems - Applications and Theory - Vol. 21*, World Scientific, s. 3-28.
- Duch W, P. Matykiewicz, J. Pestian (2007), Towards Understanding of Natural Language: Neurocognitive Inspirations. *Lecture Notes in Computer Science* **4669**, 953–962.
- Duch W, Oentaryo R.J, Pasquier M, Cognitive architectures: where do we go from here? W: *Artificial General Intelligence 2008*, Ed. by Pei Wang, Ben Goertzel, and Stan Franklin, IOS Press, pp. 122-136.
- Duch W, Pilichowski M. (2007), Experiments with computational creativity. *Neural Information Processing – Letters and Reviews* **11**, 123-133.

Duch W, Oentaryo R.J, Pasquier M, Cognitive architectures: where do we go from here?
In: Artificial General Intelligence 2008, Ed. by Pei Wang, Ben Goertzel, and Stan
Franklin, IOS Press, pp. 122-136.

Edelman G.M. (1993), Neural Darwinism: Selection and reentrant signaling in higher
brain function. *Neuron* **10**, s. 115-125.

Edelman G.M. (1999), Przenikliwe powietrze, jasny ogień. O materii umysłu.
Warszawa: PIW 1999

Feigenbaum E.A, Some Challenges and Grand Challenges for Computational
Intelligence. *J. of the ACM* **50**(1) (2003) 32–40.

Firby R.J, *Adaptive Execution in Complex Dynamic Worlds*. Ph.D. Thesis, Yale
University, 1989.

Franklin S, The LIDA architecture: Adding new modes of learning to an intelligent,
autonomous, software agent. In *Proc. of the Int. Conf. on Integrated Design and Process
Technology*. San Diego, CA: Society for Design and Process Science, 2006.

Gardner H. (1993), *Multiple intelligences: The theory in practice*. New York: Basic
Books

Goertzel B. (1997), *From Complexity to Creativity*. New York, NY: Plenum Press.

Goertzel B. (2006), *The Hidden Pattern*, BrownWalker Press.

Goertzel B, Pennachin C, Geissweiller N, Looks M, Senna A, Silva W, Heljakka A,
Lopes C. (2008), An Integrative Methodology for Teaching Embodied Non-Linguistic
Agents, Applied to Virtual Animals in Second Life, W: Artificial General Intelligence
2008, Ed. by Pei Wang, Ben Goertzel, and Stan Franklin, IOS Press, s. 161 - 175

Grąbczewski, K. N. Jankowski, Versatile and Efficient Meta-Learning Architecture:
Knowledge Representation and Management in Computational Intelligence. *IEEE
Symposium Series on Computational Intelligence* (SSCI 2007), Honolulu, HI, IEEE
Press, pp. 51-58.

- Gluck K.A, R.W. Pew (Eds.), *Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation*. Lawrence Erlbaum Associates 2005.
- Haikonen P. (2007), *Robot brains; circuits and systems for conscious machines*. Wiley.
- Hawkins J, Blakeslee S. (2004), *On intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines*. Times Books. Tł. polskie: Istota inteligencji. One Press, Helion 2006
- Hecht-Nielsen R. (2007), *Confabulation Theory: The Mechanism of Thought*. Springer.
- Hoya T. (2005), *Artificial Mind System. Kernel Memory Approach*. Springer.
- Just M.A, S. Varma (2007), The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, and Behavioral Neuroscience* **7**, s. 153-191.
- Kaelbling L.P, Littman M.L, Moore A.W. (1996), Reinforcement learning: A survey. *Journal of Artificial Intelligence Research* **4**, s.237-285.
- Kaminka G.A, C.R. Burghart (eds.) (2007), *Evaluating Architectures for Intelligence*. Technical Report WS-07-04, Menlo Park, CA: AAAI Press.
- Korner E, Matsumoto G. (2002), Cortical architecture and self-referential control for brain-like computation. *IEEE Engineering in Medicine and Biology Magazine*, **21**(5), s. 121-133.
- Laird J.E, Rosenbloom P.S, Newell A. (1987), SOAR: An architecture for general intelligence. *Artificial Intelligence* **33**, 1-64.
- Laird, J.E. (2008), *Extending the SOAR Cognitive Architecture*. W: *Frontiers in Artificial Intelligence and Applications*, Vol. 171 (Ed. by Pei Wang, Ben Goertzel, and Stan Franklin), s. 224-235.
- Lamb S. (1999), *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: J. Benjamins Publishing Co.

Langley P. (2005), An adaptive architecture for physical agents. In *Proc. of the 2005 IEEE/WIC/ACM Int. Conf. on Intelligent Agent Technology*. Compiegne, France: IEEE Computer Society Press, s. 18-25.

Langley P, Choi D. (2006), Learning recursive control programs from problem solving. *J. of Machine Learning Res.* **7**, s. 493-518.

Larvac N, Dzeroski S. (1994), *Inductive Logic Programming: Techniques and Applications*. New York: Ellis Horwood.

Manning C.D, Schütze H. (1999), *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

McClelland J.L, Rumelhart D.E. and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, MA: MIT Press.

Meeter M, Murre J.M.J. (2005), TraceLink: A model of consolidation and amnesia. *Cognitive Neuropsychology* **22** (5), s. 559-587.

Meyer D.E, Kieras D.E. (1997), A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, **104**(1), s. 3-65.

Minsky M. (1975), A Framework for Representing Knowledge. In: P.H. Winston, Ed. *The Psychology of Computer Vision*. New York: McGraw-Hill.

Minsky M. (1986), *The Society of Mind*. Simon and Schuster, New York.

Mitchell T.M, R., Keller, S. Kedar-Cabelli (1986), Explanation-based generalization: A unifying view. *Machine Learning* **1**, 47-80.

Nestor A, Kokinov B. (2004), Towards Active Vision in the DUAL Cognitive Architecture. *International Journal on Information Theories and Applications* **11**, s. 9-15.

Newell A. (1990), *Unified Theories of Cognition*: Harvard University Press.

Newell A, Simon H.A. (1963), GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and Thought*. New York: McGraw-Hill.

Newell A, Simon H.A. (1976), Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the ACM*, **19**(3), s. 113-126

Nilsson N (2005), Human-Level Artificial Intelligence? Be Serious! *The AI Magazine* 26(4), s. 68-75.

O'Reilly R.C, Braver T.S, Cohen J.D. (1999), A biologically-based computational model of working memory. W: A. Miyake i P. Shah (Eds.), *Models of Working Memory*. Cambridge University Press, pp. 375-411.

O'Reilly R.C, Munakata Y. (2000), *Computational Explorations in Cognitive Neuroscience: Understanding of the Mind by Simulating the Brain*. Cambridge, MA: MIT Press.

Panton K., C. Matuszek, D. Lenat, D. Schneider, M. Witbrock, N. Siegel, Shepard B. (2006), Common Sense Reasoning – From Cyc to Intelligent Assistant. In: Y. Cai and J. Abascal (Eds.): *Ambient Intelligence in Everyday Life*, LNAI **3864**, s. 1–31.

Pulvermuller F. (2003), *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge, UK: Cambridge University Press.

Rosbe J., R.S. Chong, Kieras D.E. (2001), Modeling with Perceptual and Memory Constraints: An EPIC-SOAR Model of a Simplified Enroute Air Traffic Control Task, SOAR Technology Inc. Report, Ann Arbor, Michigan.

Russell S, Norvig P. (2003), *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd ed.

Shanahan M.P. (2006), A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* **15**, s. 157-176.

Shapiro S.C, Rapaport W.J, Kandefer M, Johnson F.L., Goldfain A. (2007), Metacognition in SNePS, *AI Magazine* **28**, s. 17-31.

- Shastri L., Ajjanagadde V. (1993), From simple associations to systematic reasoning: A connectionist encoding of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences* **16**(3), s. 417-494.
- Sowa J.F. (1984), *Conceptual Structures*. Reading, Mass, Addison-Wesley
- Sun R, Alexandre F. (1997), *Connectionist symbolic integration*. Hillsdale, NJ: Erlbaum.
- Sun R, Merrill E., Peterson T. (2001), From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, **25**(2), 203-244.
- Sun R, Zhang X. (2004), Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research* **5**, s. 63-89.
- Szymański J, Sarnatowicz T, Duch W. (2008), Towards Avatars with Artificial Minds: Role of Semantic Memory. *Journal of Ubiquitous Computing and Intelligence* **2**, s. 1-11.
- TOSCA: A comprehensive brain-based cognitive architecture: *Biologically-Inspired Cognitive Architecture (BICA) Phase 1 Architecture Report*, DARPA-IPTO 2006.
- Turing A. (1950), Computing Machinery and Intelligence, *Mind* **49**, s. 433-460.
- Vernon D, Metta G, Sandini G. (2007), A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation* **11**(2), s. 151-180.
- Veloso M.M, Carbonell J.G. (1990), Integrating analogy into a general problem-solving architecture. In M. Zemankova & Z. Ras (Eds.), *Intelligent Systems* (pp. 29-51). Chichester, England: Ellis Horwood.
- Wallace R. (2003), *The Elements of AIML Style*, ALICE A.I. Foundation.
- Wang P. (2006), *Rigid flexibility. The Logic of Intelligence*. Springer
- Weng J, Hwang W.S. (2006), From Neural Networks to the Brain: Autonomous Mental Development. *IEEE Computational Intelligence Magazine* **1**(3), s. 15-31.