

Profesor dr hab. inż. Adam Dąbrowski
Politechnika Poznańska
Wydział Informatyki i Zarządzania
Instytut Sterowania i Inżynierii Systemów
Zakład Układów Elektronicznych i Przetwarzania Sygnałów

Poznań, 20.01.2004 r.

Recenzja wydawnicza

monografii pt.: „*Komputerowe modelowanie kompetencji językowej*”, autor:
profesor UAM dr hab. Zygmunt Vetulani

Niniejsza recenzja została przygotowana na podstawie zlecenia zawartego w piśmie od Pana Profesora dr hab. Leonarda Bolca L. Dz. IPI/NIB/4/2004 z dnia 15.01.2004 r.

Ocena trafności doboru tematyki i ogólna charakterystyka książki

Przedstawiona do oceny monografia jest poświęcona tworzeniu modeli matematycznych i wynikających z nich rozwiązań algorytmicznych problemu automatyzacji rozumienia tekstu, tj. wyszukiwania i wydobywania informacji oraz tworzenia reprezentacji wiedzy, zwłaszcza w odniesieniu do języka polskiego. Zagadnienie to jest częścią szerszej dziedziny badań nazywanej przetwarzaniem języków naturalnych (ang. *natural language processing* lub w skrócie NLP). Jest to problematyka interdyscyplinarna, łącząca wiedzę lingwistyczną z matematyką i informatyką (a zwłaszcza zagadnieniami tzw. „sztucznej inteligencji”) oraz z naukami kognitywnymi. Nabiera ona zwłaszcza w ostatnich latach coraz większego znaczenia w związku z intensywnym rozwojem systemów automatyki i robotyki wyposażonych w moduły bezpośredniej komunikacji z użytkownikiem (ang. *human-machine-interface* lub w skrócie HMI).

Od modułu HMI, który nadzoruje niższe warstwy systemu automatyki, planuje i przydziela zadania, oczekuje się bezbłędnego działania w czasie rzeczywistym, co z jednej strony wymaga wysokiej wydajności obliczeniowej, a z drugiej – doskonałej emulacji kompetencji językowej. Oceniana książka jest poświęcona drugiemu z wymienionych zagadnień. Podjęta w niej problematykę oceniam jako w pełni aktualną i bardzo interesującą, o trudnym do przecenienia znaczeniu poznawczym i aplikacyjnym w nowoczesnych systemach automatyki i robotyki. Praca ta, biorąc zwłaszcza pod uwagę aspekt rozważań skoncentrowanych na zastosowaniach do języka polskiego, ma charakter pionierski.

Należy podkreślić znaczny wkład Autora w rozwój omawianej dziedziny. W książce są zaprezentowane wyniki Jego wieloletnich badań naukowych, rozproszone dotychczas w licznych specjalistycznych i stosunkowo hermetycznych publikacjach. Zebranie ich i opublikowanie w formie niniejszej monografii jest bardzo cenną inicjatywą, która nie tylko porządkuje dotychczasową wiedzę, ale uogólnia ją i popularyzuje, wytyczając jednocześnie kierunki jej dalszego rozwoju.

Ocena merytoryczna książki

Celem recenzowanej monografii jest – jak to formułuje jej Autor w Przedśłowiu – „wprowadzenie Czytelnika w podstawy jednej z najstarszych a jednocześnie ciągle nabierającej znaczenia dziedzin informatyki, jaką jest przetwarzanie języka naturalnego (NLP)”. Moim zdaniem Autor realizuje cel znacznie ambitniejszy, polegający nie tylko na zaznajomieniu z podstawami NLP, lecz także na opisanie zaawansowanego stanu obecnego tego nowatorskiego obszaru badawczego, zarysowanego na tle rozwoju takich dziedzin pokrewnych jak nauka o języku, czy sztuczna inteligencja i to zarówno w warstwie teoretycznej abstrakcyjnych modeli matematycznych jak i w warstwie aplikacyjnej konkretnych realizacji algorytmicznych.

Obszerny tekst książki zawierający 247 stron jest podzielony na dziewięć części, uzupełnionych przedśłowiem, bibliografią pięcioma dodatkami i ponadto starannie opracowanym skorowidzem. Po wprowadzeniu w części I, zawierającym wyjaśnienie pojęcia kompetencji językowej oraz informacje o związkach analizowanych zagadnień z teorią języka i informatyką Autor zaznajamia Czytelnika w części II z wybranymi metodami formalnego opisu języka naturalnego, w tym z opisem składniowym za pomocą systemu reguł i z teorią gramatyk formalnych Chomsky'ego. Przedstawia także algorytmny parsujący, tzn. tworzące drzewa składniowe, wyznaczające segmentację zdań, a następnie podaje ogólny, predykatowo-argumentowy model zdania. Podstawy teoretyczne opisane w części II są przez Autora wykorzystane w części III do stworzenia formalnego modelu opisu składniowego języka polskiego, przydatnego do przetwarzania za pomocą komputera. Część IV jest poświęcona omówieniu tendencji rozwojowych budowy tzw. „leksykonu-gramatyki” dla języka polskiego. W części V Autor omawia problem rozwoju uproszczonych strategii poruszania się po przestrzeni przeszukiwań (tj. przeszukiwania heurystycznego) w celu zmniejszenia złożoności obliczeniowej algorytmów parsingu. Rozwijając omawiane zagadnienia, Autor koncentruje się w części VI na tworzeniu baz reprezentacji wiedzy na podstawie analizy semantycznej tekstu. Część VII zawiera prezentację przykładowego, ilustrującego dotychczasowe rozważania, schematu organizacyjnego procesu analizy zdań. Rozwiązanie problemu parsingu w przypadku gramatyk szyku swobodnego jest podane w części VIII. Na zakończenie Autor zajmuje się w części IX wyszukiwaniem i wydobywaniem informacji na podstawie analizy tekstu dokumentów.

Podział monografii na omówione powyżej części, a także rozdziały i podrozdziały oceniam jako dokonany w pełni prawidłowo, tj. wyczerpujący zakres tematyczny prezentowanych badań i dobrze porządkujący zawarte w nich informacje. Tekst książki jest jasny i ścisły, choć napisany stosunkowo hermetycznym językiem, wymagającym od czytelnika odpowiedniego przygotowania. Należy podkreślić, że jest opracowany bardzo starannie zarówno pod względem edytorskim jak i merytorycznymi, praktycznie nie zawiera usterek typograficznych (informacje o nielicznych, które zauważyłem, przekazałem Autorowi).

Nie mam żadnych merytorycznych uwag krytycznych. Polemizować mógłbym jedynie z używanym przez Autora pojęciem emulacja w sensie modelowania.

Reasumując, moim zdaniem oceniana książka w pełni zasługuje na jej opublikowanie. Będzie z pewnością kolejną cenną pozycją z zakresu informatyki i sztucznej inteligencji wydaną przez Wydawnictwo Exit.