

Tomasz KAPŁON

Politechnika Wroclawska, Instytut Informatyki, Automatyki i Robotyki

## WIZUALIZACJA ZNACZENIA TREŚCI PRZEKAZU FORMUŁOWANEGO W JĘZYKU NATURALNYM

**Streszczenie.** W artykule przedstawiono metodę: odkrywania znaczenia treści przekazu oraz tworzenia obrazu semantycznego takiego przekazu, organizację bazy wiedzy systemu generującego obraz, reprezentujący znaczenie treści, sposób umieszczania obiektów na scenie oraz krótką dyskusję na temat testowania zgodności wygenerowanego obrazu z wyobrażeniem obserwatora.

**Słowa kluczowe:** przetwarzanie języka naturalnego, przetwarzanie tekstu w obraz, reprezentacja wiedzy

## CONTENT VISUALIZATION OF TEXT COMPOSED IN NATURAL LANGUAGE

**Summary.** The method of content meaning discovering and semantic image creating, structure of knowledge base of system generating image representing meaning content, the method of placing object on the 3D scene and short discussion about testing correspondence of generated image with imagination of observer is presented in the paper.

**Keywords:** natural language processing, text to image transformation, knowledge representation

### 1. Wprowadzenie

Obraz to więcej niż tysiąc słów. Dla większości ludzi to powiedzenie jest oczywiste i zrozumiałe. Gdy widzimy obraz lub serię obrazów, to jesteśmy w stanie pełniej zrozumieć ich znaczenie, niż wtedy kiedy mamy do dyspozycji jedynie tekst. W przypadku, w którym tekst jest napisany w języku dla nas nienaturalnym, problem ze zrozumieniem jego znaczenia staje się trudniejszy. Kiedy jednak to, co zawarte jest w tekście zobaczymy w postaci obrazu,

nagle całość staje się oczywista i rozumiała lub przynajmniej bardziej rozumiała, często dokładnie w ten sam sposób, w jaki autor tekstu chciał, abyśmy to zobaczyli. Rzecz jasna, że nasze rozumienie jest silnie zależne od wiedzy, którą posiadamy w danym temacie lub też od posiadanych doświadczeń, szczególnie w przypadku rozumienia pojęć abstrakcyjnych. Niemniej jednak prezentacja informacji w postaci obrazów jest formą bardziej komunikatywną, zwłaszcza w przypadku konieczności komunikacji osób o różnym pochodzeniu, różnie bogatym aparacie pojęciowym czy różnej umiejętności posługiwania się językami.

Komunikatywne przekazywanie informacji szerokiej rzeszy ludzi, dla której nie istnieje uniwersalny język komunikacji wymaga istnienia systemów komunikacji obrazowej. Podstawową cechą takich systemów powinna być możliwość interpretowania treści komunikatów formułowanych w językach naturalnych i prezentacji znaczenia tych komunikatów w postaci obrazów. Aby tego dokonać, niezbędne są mechanizmy oparte na modelach formalnych przetwarzania języka w postaci tekstu, pozwalających na odkrywanie znaczenia treści przekazu. Systemy obrazowania znaczenia tekstu znane są w literaturze jako systemy text-to-scene. Systemy tego typu mają, w zależności od rozwiązania, zdolność generowania obrazu na podstawie tekstu szczegółowo opisującego obiekty i ich położenie: NALIG [1], SPRINT[2] czy WordEye[3], na podstawie zestawu słów kluczowych [4, 5], jednak posiadają taką zaletę, że operują na tekstach formułowanych w języku naturalnym. Brakuje systemów poprawnie interpretujących treść niejako niezależnie od sposobu sformułowania tekstu, nawet niepoprawnego składniowo, jednak poprawnego semantycznie. Systemem spełniającym, po części, to zadanie jest propozycja Goldberga [6], z tym zastrzeżeniem, że obraz konstruowany jest na podstawie wyszukanych w Internecie zdjęć i grafiki, co nie spełnia jednego z dwóch założeń proponowanego tutaj systemu, czyli obrazu sceny, spójnego wizualnie.

Treść artykułu stanowi rozwinięcie rozwiązania prezentowanego w [7] i zawiera pełniejszą prezentację odkrywania znaczenia treści (tworzenia obrazu semantycznego przekazu), organizację bazy wiedzy, sposób umieszczania obiektów na scenie oraz krótką dyskusję na temat testowania zgodności wygenerowanego obrazu z wyobrażeniem obserwatora.

## **2. Odkrywanie znaczenia treści**

### **2.1. Scenariusz**

Przekaz wprowadzony do systemu poddawany jest systematycznym analizom syntaktycznej i semantycznej. Teoretycznie istotne jest jedynie znaczenie przekazu, a co za tym idzie, poprawność syntaktyczna powinna nie mieć, w przypadku odkrywania znaczenia treści przekazu, żadnego lub przynajmniej kluczowego znaczenia. Niemniej jednak analiza seman-

tyczna z użyciem formalnego modelu obrazowania semantyki wymaga, aby zdania przekazu były syntaktycznie poprawne. Niestety wykorzystywany model zmodyfikowanej gramatyki łączy podczas podziału zdań na grupy słów sprawdza ich poprawność składniową. W przypadku braku możliwości dopasowania łączy zdanie, oznaczane jest jako niepoprawne nie jest dzielone na grupy słów i tym samym nie jest analizowana semantyka zdania. Dlatego też każde zdanie w scenariuszu musi być poprawne syntaktycznie, aby możliwe było określenie znaczenia przekazu i jego graficzna prezentacja. Poniżej przedstawiona zostanie analiza syntaktyczna scenariusza. Pokazany zostanie podział na grupy słów oraz określenie obrazu semantycznego trzech zdań. Dekonstrukcja pozostałych zostanie ujęta w kilku wyrażeniach symbolicznych, reprezentujących całość scenariusza. Będą to wyrażenia  $A^{*+}$  – reprezentujące zdania składające się na scenariusz, tak zwaną reprezentację syntaktyczną,  $B^{*+}$  – reprezentacja semantyczna (zestaw obrazów semantycznych zdań scenariusza) oraz  $U^{*+}$  kojarzące ze sobą reprezentację syntaktyczną i semantyczną.

Rozpatrzmy następujący, prosty scenariusz.

*A grey block is put on a white basis.* (1)

*Ten centimeters on the right is set a white block.* (2)

*Some sphere is blue and hangs over a purple pyramid.* (3)

*The sphere has a radius of three centimeters.* (4)

*The purple pyramid stands three centimeters over the basis.* (5)

*A red cylinder stands on the white block.* (6)

## 2.2. Analiza syntaktyczna

W przypadku omawianego systemu analiza syntaktyczna odbywa się na podstawie modelu formalnego zmodyfikowanej gramatyki łączy [8] oraz na modelu formalnego obrazowej reprezentacji semantyki zdań [9, 10]. Scenariusz składający się z dowolnej liczby zdań opisuje scenę. Rozważmy zdania (1) i (2), aby pokazać proces gromadzenia wiedzy.

*A gray block is put on a white basis.* (1)

Podział na grupy słów zdania (1) jest następujący (rdzeń grupy wytłuszczony):

*a gray **block*** –  $d_1 d_2 d_3 \leftrightarrow a_1$

*is **put** on* –  $d_4 d_5 d_6 \leftrightarrow a_2$

*the white **basis*** –  $d_7 d_8 d_9 \leftrightarrow a_3$

$s_1 = a_1 a_2 a_3$

Symbole  $d_j$  reprezentują wyrazy,  $a_i$  grupy słów i  $s_n$  zdanie. Każde zdanie można przedstawić w postaci grafu, który jest równoważny wyrażeniu [10]. Poszczególne grupy i zdania gromadzone są w bazie wiedzy, w postaci wyrażen symbolicznych. Dla każdego zdania tworzone

jest wyrażenie reprezentujące jego składnię  $A^+(s_n)$ . Dla zdania (1) wyrażenie ma następującą postać:

$$A^+(s_1) = {}^0(a_1 {}^1(z_1 s_1 a_2 {}^2(z_1 s_1 a_3 {}^3(z_3 s_1 a_1)^3)^2)^1)^0 = A'^+ \quad w_1$$

Symbole  $z_1$  stanowią łączniki pomiędzy kolejnymi grupami słów w zdaniu, natomiast symbol  $z_3$  określa połączenie pomiędzy grupami ostatnią i pierwszą.

Wyrażenie  $w_1$  jest równoważne wyrażeniu reprezentującemu składnię wszystkich zdań znajdujących się aktualnie w bazie wiedzy  $A'^+$ . W bazie wiedzy gromadzona jest również reprezentacja grup słów w postaci wyrażenia  $D'^+$ . Jest to konieczne, aby spełnić jeden z warunków poprawnej reprezentacji wiedzy, czyli możliwość jej odtworzenia.

Poprawność syntaktyczna zdań weryfikowana jest na podstawie zmodyfikowanej gramatyki łączy. Cechą modelu przedstawionego w [8] jest właśnie kontrola poprawności syntaktycznej zdań.

Zdanie (2)

*Ten centimeters on the right is set a white block.*

dzielone jest w następujący sposób:

*a white block* –  $d_0 d_7 d_3 \leftrightarrow a_4$

*is set* –  $d_4 d_{13} \leftrightarrow a_5$

*ten centimeters on the right* –  $d_9 d_{10} d_6 d_{11} d_{12} \leftrightarrow a_6$

Wyrażenie reprezentujące składnię zdania (2) ma następującą postać:

$$A^+(s_2) = {}^0(a_4 {}^1(z_1 s_2 a_5 {}^2(z_1 s_2 a_6 {}^3(z_3 s_2 a_4)^3)^2)^1)^0 \quad w_2$$

Wyrażenie  $A'^+$  ma postać:

$$A'^+ = {}^0(a_1 {}^1(z_1 s_1 a_2 {}^2(z_1 s_1 a_3 {}^3(z_3 s_1 a_1)^3)^2)^1, a_4 {}^1(z_1 s_2 a_5 {}^2(z_1 s_2 a_6 {}^3(z_3 s_2 a_4)^3)^2)^1)^0 \quad w_3$$

Analizując kolejne zdania otrzymamy wyrażenie  $w_4$  reprezentujące całość tekstu:

$$A'^+ = {}^0(a_1 {}^1(z_1 s_1 a_2 {}^2(z_1 s_1 a_3 {}^3(z_3 s_1 a_1, z_3 s_6 a_9 {}^4(z_3 s_4 a_7, z_1 s_6 a_{12} {}^5(z_1 s_6 a_3)^5, a_7 {}^5(z_3 s_3 a_7, z_3 s_5 a_7, z_3 s_4 a_8 {}^6(z_1 s_4 a_9)^6)^5)^4)^3)^2)^1, a_4 {}^1(z_1 s_2 a_5, z_1 s_7 a_{10} {}^3(z_1 s_7 a_4)^3)^2, a_5 {}^2(z_1 s_2 a_6 {}^3(z_3 s_2 a_4)^3)^2)^1)^0$$

Wyrażenie to można przedstawić w postaci równoważnego grafu. W bazie wiedzy natomiast każdemu słowu (symbolowi  $d_j$ ) przypisany jest obiekt (najczęściej opis jego tworzenia), który może zostać umieszczony na scenie, jako graficzne przedstawienie znaczenia pojęcia lub pojęć (rys. 1).

### 2.3. Analiza semantyczna

Efektom analizy semantycznej zdania jest stworzenie obrazu semantycznego tekstu na poziomie kategorii semantycznych. Polega to na nadaniu rdzeniowi grupy cechy semantycznej (symbol  $b_m$ ) i na tej podstawie określenie kategorii semantycznej grupy [9]. Dla zdania (1) wygląda to następująco.

*a gray block*  $\leftrightarrow a_1$  – OBJECT  $\leftrightarrow b_{n1}$

*is put on*  $\leftrightarrow a_2$  – ACTION  $\leftrightarrow b_{v9}$

*the white basis*  $\leftrightarrow a_3$  – LOCATION  $\leftrightarrow b_{n3}$

Następnie, dla każdego zdania, określa się semantykę  $c_m$  oraz strukturę semantyczną zdania na poziomie cech semantycznych  $B^+(c_m)$ . Następnie tworzy się wyrażenie  $U^+(c_m)$ , określające zależność pomiędzy strukturami syntaktyczną  $A^+(c_m)$  i semantyczną  $B^+(c_m)$ . Sposób wyznaczania semantyki  $c_m$  dla zdania polega na sprawdzaniu istnienia w bazie wiedzy wyrażenia  $B^+(c_m)$  identycznego jak fragment wyrażenia  $B^+$ , w którym zgromadzona jest wiedza na temat struktury semantycznej na poziomie cech semantycznych wszystkich zdań, wprowadzonych do bazy wiedzy. Jeżeli w bazie istnieje obraz semantyczny identyczny jak ten, którego szukamy, nadawana jest mu określona już semantyka, w przeciwnym wypadku ustalana jest nowa. Dla zdań scenariusza wyrażenie  $B^+$  (reprezentujące ogólny obraz semantyczny przekazu) wygląda następująco:

$$B^+ = {}^0(b_{n3} {}^1(u_1 c^+ {}_1 b_{v9} {}^2(u_2 c^+ {}_1 b_{n3} {}^3(u_0 c^+ {}_1 b_{n3}, u_0 c^+ {}_3 b_{n3} {}^4(u_1 c^+ {}_7 b_{n3}, u_1 c^+ {}_3 b_{n11} {}^5(u_2 c^+ {}_3 b_{n3})^5, b_{n3} {}^5(u_0 c^+ {}_7 b_{n3}, u_0 c^+ {}_5 b_{n3}, u_0 c^+ {}_7 b_{v11} {}^6(u_2 c^+ {}_7 b_{n11})^6)^5)^4)^3)^2)^1, b_{n3} {}^1(u_1 c^+ {}_2 b_{v9}, u_1 c^+ {}_6 b_{v11} {}^3(u_2 c^+ {}_6 b_{n3})^3)^2, b_{v9} {}^2(u_2 c^+ {}_2 b_{n11} {}^3(u_0 c^+ {}_2 b_{n3})^3)^2)^1)^0$$

W ten sposób w bazie wiedzy gromadzona jest wiedza na temat struktury składniowej i struktury znaczeniowej zdań tworzących scenariusz. Operacje na tych strukturach pozwalają na interpretację znaczenia poszczególnych grup słów i prezentację tego znaczenia w postaci obrazu.

W efekcie mamy do dyspozycji wiedzę na różnych poziomach abstrakcji, gdzie najważniejsze jest posiadanie obrazu semantycznego przekazu, czyli to, co należało wydobyć z jego treści i co pozwala na pokazanie tego znaczenia w postaci obrazu, oczywiście w ramach zgromadzonej wiedzy.

### 3. Generowanie obrazu sceny

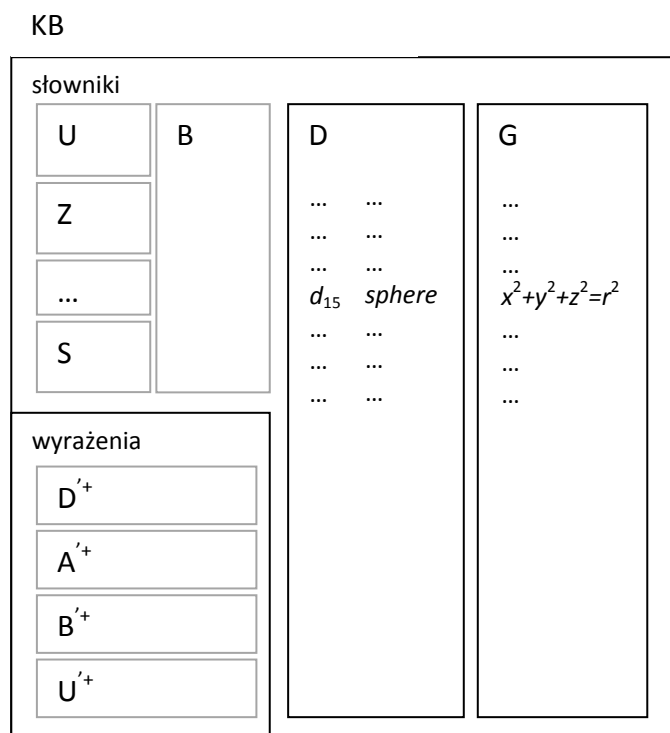
#### 3.1. Organizacja bazy wiedzy

Baza wiedzy dla prezentowanego systemu jest bazą systemu obrazowania semantyki zdań (rys. 1), która uzupełniona jest o połączenia pomiędzy tabelą  $D$  a tabelą  $G$ , w których umieszczone są podstawowe obiekty graficzne. W szczególności w tabeli  $G$  zgromadzone są opisy tworzenia podstawowych obiektów graficznych, w tym wszystkie parametry umożliwiające modyfikowanie obiektu.

Modyfikacje dokonywane są na podstawie treści przekazu. Możliwych jest kilka przypadków wykonania modyfikacji:

- gdy jawnie podane są konkretne wartości parametrów (np. wysokość),

- kiedy podane są zależności pomiędzy obiektami (np. auto ma wagę dorosłego słonia),
- kiedy modyfikacja parametru podana jest w postaci rozmytej (piramida jest wyższa od namiotu),
- kiedy dokonywana jest na podstawie proporcji pomiędzy obiektami, które wynikają z ich definicji zawartej w tabeli  $G$ , jeśli w tekście nie zostały określone żadne informacje dotyczące modyfikacji obiektu; na podstawie zdefiniowanych proporcji konieczna jest korekta dla zachowania spójności wizualnej sceny.



Rys. 1. Schemat bazy wiedzy

Fig. 1. Knowledge base scheme

### 3.2. Umieszczenie obiektów na scenie

Umieszczania obiektów na scenie, w ogólnej postaci, odbywa się kilkuetapowo:

- 1) umieszczenie obiektów w pozycjach jednoznacznie określonych – przypadek rzadki biorąc pod uwagę, że podawanie konkretnej lokacji obiektów za pomocą współrzędnych występuje raczej w dokumentacji technicznej niż w swobodnym przekazie informacji, opartym na intuicyjnym założeniu o wspólnej wiedzy o temacie rozmowy,
- 2) umieszczany jest w lokacji domyślnej, jeżeli obiekt nie ma pozycji określonej bezwzględnie, w najprostszym przypadku jest początek układu współrzędnych sceny; w bardziej zaawansowanej wersji można sobie wyobrazić istnienie pewnej wiedzy na temat sytuacji opisanej w przekazie i na tej podstawie określenie różnych (dla poszczególnych obiektów) miejsc domyślnego ustawienia,

- 3) pozycjonowanie względne opierające się na miarach jednoznacznie określone; jest to korekta ustawienia domyślnego,
- 4) modyfikacja parametrów obiektów na podstawie pojęcia rozmytego oraz ustalenie spójności wizualnej sceny.

Scenę uznajemy za spójną wizualnie wtedy, gdy wszystkie umieszczone na niej obiekty pozostają wobec siebie w takiej relacji, która jest zbieżna z prawami świata, dla którego ugruntowane zostały pojęcia i ich związki, użyte w przekazie.

### 3.3. Testowanie poprawności wykonania

Problem poprawności wygenerowanej sceny (w ogólnym znaczeniu) jest trudny do określenia. Poprawność jest niepodważalna, jeżeli mamy do czynienia z przekazem, w którym wszystkie obiekty mają ustaloną, bezwzględną (jednoznaczną) pozycję na scenie. Wtedy, sprawdzenie poprawnego odwzorowania nie jest trudne i nie pozostawia wątpliwości. W przypadku, w którym pojęcia lub ich wzajemne zależności mogą być różnie interpretowane (wyobrażane) przez odbiorców, nie ma jednoznacznego sposobu na autonomiczny test poprawności wygenerowania sceny. Ocena taka zależna jest bowiem od subiektywnego odbioru – porównania tego, co obserwator widzi z tym, co sam sobie konstruuje (wyobraża), w ramach własnego obrazowego aparatu pojęciowego. Można jednak powiedzieć, że poprawność wygenerowania sceny jest wprost proporcjonalna do wiedzy, która zawarta jest w bazie wiedzy. Wniosek ten jest w pewnym sensie efektem obserwacji działania opisywanego systemu, jednak w większej mierze bazuje on na obserwacji osób wzbogacających swoją wiedzę dziedzinową i ich zdolność do wyciągania bardziej złożonych i bardziej poprawnych wniosków, w ramach rozwiązywanych problemów z danej dziedziny.

## 4. Podsumowanie

Przedstawiono zagadnienie odkrywania znaczenia treści przekazu oraz tworzenia obrazu semantycznego takiego przekazu, dzięki czemu możliwe jest prezentowanie znaczenia przekazu w postaci sceny generowanej z użyciem obiektów graficznych, których definicja dostępna jest w bazie wiedzy. Zaproponowano sposób rozmieszczania obiektów na scenie, w którym trzy pierwsze punkty zostały wstępnie zaimplementowane. Punkt czwarty, czyli modyfikacja sceny na podstawie pojęć rozmytych, wymaga opracowania procedur analizy treści przekazu. Jednocześnie niezbędne jest ustalenie rodzaju opisu dla obiektów złożonych, których opis w postaci równań jest niemożliwy lub trudny (np. hipopotam). Pożądanym efektem

jest bowiem możliwość kreacji sceny bądź serii scen 3D przez system złożony z modułu analizy treści przekazu oraz modułu generowania scen.

## BIBLIOGRAFIA

1. Adorni G., Manzo M. D., Giunchiglia F.: Natural language driver image generation. [in:] Proc. COLING, 1984.
2. Yamada A., Yamamoto T., Ikeda H., Nishida T., Doshita S.: Reconstructing spatial image from natural language texts, [in:] Proc. COLING, vol. 4, 1992.
3. Coyne B., Sproat R.: WordsEye: An automatic text-to-scene conversion system. [in:] Proc. SIGGRAPH, 2001.
4. Mihalcea R., Leong B.: Toward Communication Simple Sentences Using Pictorial Representations, [in:] Proc. Conf Association for Machine Translation in the Americas (AMTA), 2006.
5. Lu R., Zhang S.: Automatic Generation of Computer Animation: Using AI for Movie Animation. Lecture Notes in AI, Vol. 2160. Springer-Verlag, Berlin 2001.
6. Goldberg A. B., Xiaojin Z., Dyer C. R., Eldway M., Heng L.: Easy as ABC? Facilitating Pictorial Communication via Semantically Enhanced Layout. In Twelfth Conference on Computational Natural Language Learning (CoNLL), 2008.
7. Kapłon T.: Graficzna interpretacja znaczenia tekstu. *Studia Informatica*, Vol. 31, No. 2A, Silesian University of Technology Press, Gliwice 2010, s. 163÷172.
8. Mierzwa J.: Model formalny komputerowej dekompozycji zdań języka naturalnego na grupy słów do celów wnioskowania przez komputer. Raport seria PRE 39/2001 (praca doktorska), Oficyna Wydawnicza PWR, 2001.
9. Kapłon T.: Model formalny obrazowej reprezentacji semantyki zdań języka naturalnego do wnioskowania przez computer. Raport seria PRE 30/2003 (praca doktorska), Oficyna Wydawnicza PWR, 2003.
10. Kapłon T.: Model formalny obrazowej reprezentacji semantyki zdań języka naturalnego. Bazy danych: rozwój metod i technologii: bezpieczeństwo, wybrane technologie i zastosowania, praca zbiorowa (pod red.) Kozielski S. [i in.], Wydawnictwa Komunikacji i Łączności, Warszawa 2008.



Recenzenci: Dr inż. Małgorzata Bach  
Dr hab. inż. Krzysztof Goczyła, prof. Pol. Gdańskiej

Wpłynęło do Redakcji 31 stycznia 2011 r.

### **Abstract**

The method of content meaning discovering and semantic image creating, structure of knowledge base of system generating image representing meaning content, the method of placing object on the 3D scene and short discussion about testing correspondence of generated image with imagination of observer is presented in the paper.

### **Adres**

Tomasz KAPŁON: Politechnika Wrocławska, Instytut Informatyki, Automatyki i Robotyki  
ul. Janiszewskiego 11/17, 50-372 Wrocław, Polska, tomasz.kaplon@pwr.wroc.pl.