AUTOMATYKA • 2008 • Tom 12 • Zeszyt 2

Ryszard Tadeusiewicz*, Marek R. Ogiela*

Nowy element w instrumentarium inteligencji obliczeniowej – automatyczne rozumienie obrazów

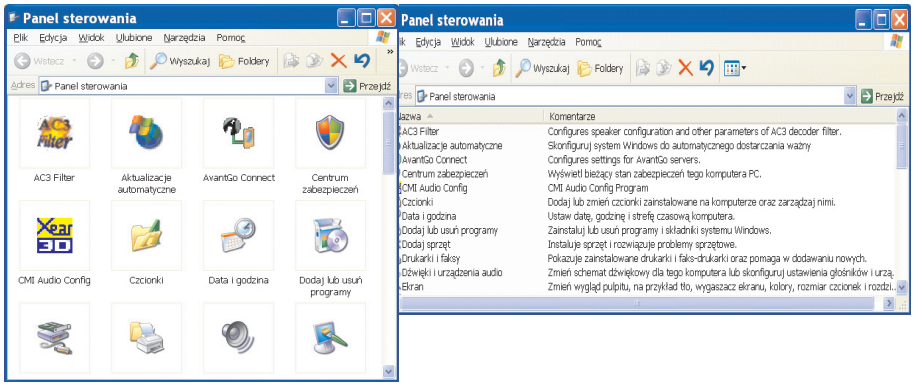
1. Wprowadzenie

Zafascynowani różnymi aspektami rozwoju i upowszechnienia informatyki i telekomunikacji, którego najbardziej dostrzegalnym przejawem jest błyskawiczny rozwój Internetu, nie dostrzegamy jednego zjawiska, jakie temu rozwojowi towarzyszy. Zjawiskiem tym jest postępująca migracja od cywilizacji pisma w kierunku cywilizacji obrazu, albo wyrażając to samo inaczej – od informacji tekstowych ku informacjom multimedialnym.

Zwiastunem tego procesu było całkowite wyparcie powszechnie używanych (niegdyś) interfejsów alfanumerycznych przez interfejsy graficzne typu Windows [1]. Jeszcze niedawno GUI (*Graphical User Interface*) był utożsamiany z aplikacjami przeznaczonymi dla najmniej wykwalifikowanych użytkowników informatyki. Profesjonaliści używali komend tekstowych, które potrafili zresztą splatać w niezwykle wyrafinowane polecenia. Celowali w tym zwłaszcza administratorzy systemów opartych na Uniksie (a potem także Linuksie). Obecnie także te enklawy myślenia i działania z wykorzystaniem wyrafinowanych kodów zostały całkowicie zdominowane przez obrazki, ikony, operacje wykonywane myszką. Obrazkowy język komunikacji z komputerem przestał być antytezą profesjonalizmu, ponieważ także profesjonaliści dostrzegli, że dzięki stosowaniu interfejsów graficznych mogą łatwiej zapanować nad złożonymi systemami i skomplikowanymi procesami [2]. Co więcej, wszyscy przekonali się, że polecenia wydawane metodą klikania myszką są nie tylko szybsze i wygodniejsze (z punktu widzenia człowieka) niż wpisywanie haseł czy tekstowych poleceń, ale wiążą się także z mniejszym prawdopodobieństwem popełnienia błędu (rys. 1).

Zalety interfejsu graficznego doceniono także w kontekście sterowania wieloma innymi systemami technicznymi, dlatego obsługiwane dotykowo ikony zamiast napisów pojawiły się na wyświetlaczach telefonów komórkowych, na panelach sterujących różnych maszyn i urządzeń, w kokpitach nowoczesnych samochodów i samolotów itd.

* Katedra Automatyki, Akademia Górniczo-Hutnicza w Krakowie



Rys. 1. Te same informacje przedstawione w formie graficznej i w formie tekstowej ujawniają zalety graficznej prezentacji

Wspomniany wyżej GUI to przysłowiowy wierzchołek góry lodowej. Człowiek jest wzrokowcem i ponad 90% informacji pozyskuje za pomocą percepcji i analizy różnych obrazów. Dlatego w zakresie przekazywania ludziom dowolnych informacji ich prezentacja graficzna ma zdecydowaną przewagę nad wszelkimi innymi sposobami i formami prezentacji. Oglądając obraz, można łatwiej przyswoić sobie liczne i złożone informacje, skuteczniej dostrzega się powiązania pomiędzy różnymi (na pozór) faktami, szybciej i poprawniej dociera się do rzeczywistego znaczenia prezentowanych danych oraz łatwiej można przeprowadzić jakieś złożone rozumowanie, które się do tych danych odnosi.

2. Zalety obrazów jako nośników informacji i ich konsekwencje

Zalety graficznej prezentacji informacji najwcześniej odkryli specjaliści od nauczania, wśród których bardzo popularne jest powiedzenie *jeden obraz to więcej, niż tysiąc słów*. Aktualnie graficzne prezentacje stają się coraz popularniejsze przy przedstawianiu różnych wyników obliczeń (na przykład ekonomicznych) w ramach tzw. wizualizacji danych [3]. Graficzna forma prezentacji była, jest i będzie podstawową formą prezentacji w odniesieniu do danych o terenie (plany geodezyjne, mapy geograficzne, zdjęcia lotnicze i satelitarne itp.).

Zasygnalizowane wyżej zagadnienia skłaniają do sformułowania dwóch pytań:

1) Dlaczego mimo tak licznych zalet obrazów jako nośnika informacji ludzie nie rozwinęli na szerszą skalę metod komunikacji obrazowej?
2) Co powoduje, że w miarę postępu informatyzacji proporcje między informacjami przedstawianymi w formie tekstowej a informacjami przedstawianymi w formie obrazowej ustawicznie się zmieniają na korzyść tych drugich?

Odpowiedź na oba pytania opiera się na konstatacji jednego, powszechnie znanego faktu. Otóż człowiek nie posiada efektywnego narządu, z pomocą którego mógłby szybko i wygodnie wytwarzać obrazy, zawierające treści, które chciałby zakomunikować innym

ludziom. Gdybyśmy potrafili świadomie i celowo wytwarzać na jakiejś powierzchni wchodzącej w skład naszego ciała dowolne pomyślane obrazy równie szybko, łatwo i naturalnie, jak generujemy sygnał mowy, to nasza cywilizacja prawdopodobnie wyglądałaby dziś całkiem inaczej. Jest jednak inaczej. To z pomocą mowy potrafimy wrażać nasze myśli, dlatego utrwalanie tych myśli przybiera postać pisma, czyli mowy utrwalonej, zaś obrazy, gdy już zdecydujemy się je sporządzać, wytwarzamy z dużym nakładem pracy, nie zawsze opłacalnej. W dodatku, żeby ręcznie wytwarzać obrazy, potrzebny jest pewien talent oraz zasoby określonych narzędzi – w sumie więc mowa (oraz pismo) są przy przekazywaniu i gromadzeniu informacji regułą, a obrazy stanowią wygodny, miły, chętnie akceptowany przez odbiorcę dodatek do komunikatów językowych.

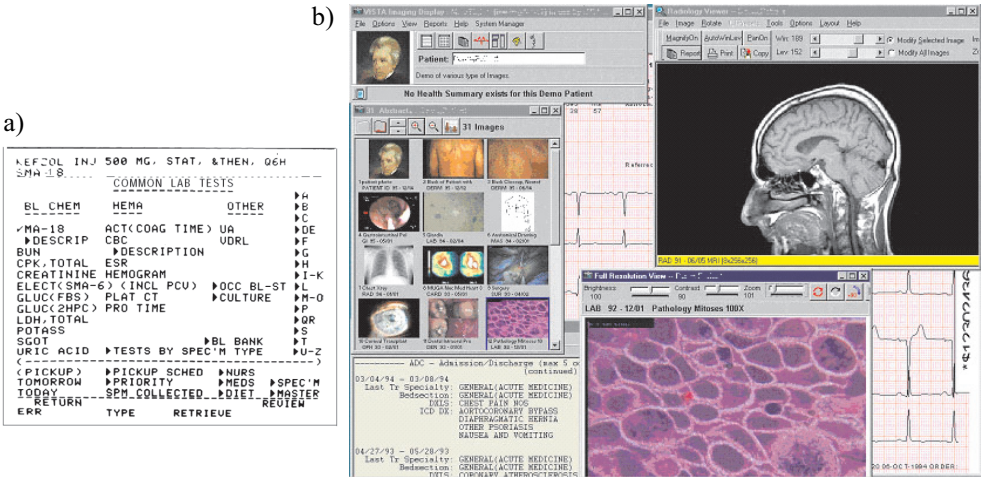
Tak było przez całe stulecia – aż do wynalezienia komputera oraz grafiki komputerowej jako dziedziny zastosowań. Rozwój tej ostatniej w połączeniu z powszechną dostępnością urządzeń do cyfrowej rejestracji obrazów obiektów rzeczywistych (cyfrowe aparaty fotograficzne i kamery wideo, rozmaite skanery i urządzenia do digitalizacji) spowodowały łącznie, że dziś wytworzenie obrazu komputerowego wymaga bardzo niewiele wysiłku. Czasem wystarczy tylko prosta decyzja, jak w przypadku konwersji tabeli liczb do postaci wykresu dwu- lub nawet trójwymiarowego, co kiedyś wymagało wielu godzin wytężonej pracy, a dziś otrzymywane jest w ułamku sekundy po pojedynczym kliknięciu myszką. W dodatku dostępne są liczne i bardzo wygodne narzędzia informatyczne, służące do tworzenia i przetwarzania obrazów, a używanie tych narzędzi nie wymaga posiadania specjalnego talentu, tylko odrobiny wprawy i cierpliwości.

Łatwość rejestrowania obrazów obiektów rzeczywistych oraz wygoda tworzenia obrazów tworów pomyślanych, powoduje, że w różnych zasobach informacyjnych coraz częściej występują głównie obrazy, a udział tekstu bezustannie się zmniejsza. Dobrym przykładem może tu być cyfrowo zarejestrowana dokumentacja pacjenta w nowoczesnym szpitalu, która dawniej miała charakter głównie tekstowy, a dziś głównie obrazowy (rys. 2).

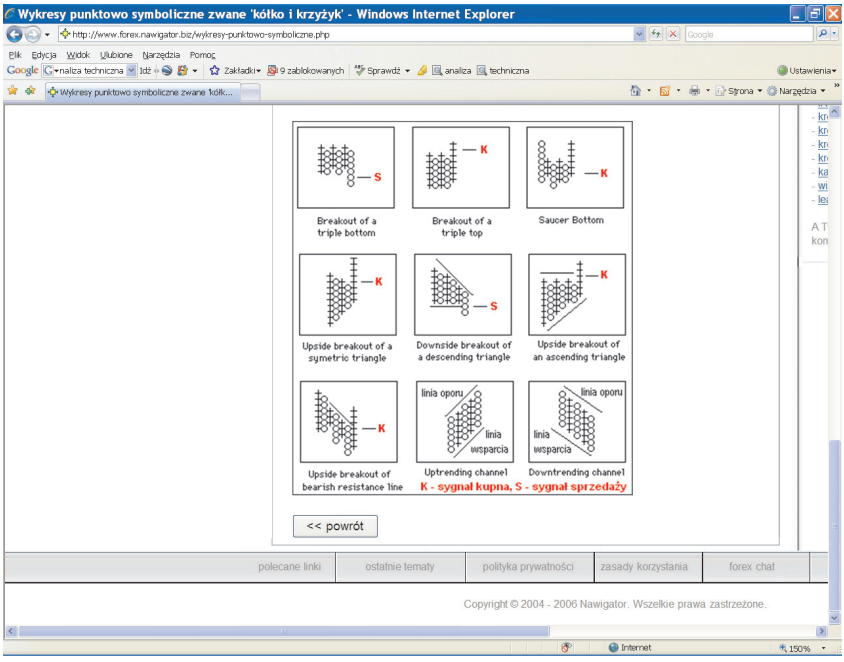
Dawniej dokumentacja taka miała głównie postać określonych zapisów tekstowych, w których mieściło się wszystko: dane personalne, opis przypadku, opisy badań, postawiona diagnoza, zastosowana terapia oraz osiągnięte wyniki. Dziś opisy tekstowe także występują, ale w reprezentacji komputerowej zajmują zaledwie kilkadziesiąt kilobajtów. Natomiast w dokumentacji dominują różne formy cyfrowych zobrazowań medycznych – zapisy z tomografu, ultrasonografu, rezonansu magnetycznego, badań PET i wielu innych. Zapisy te zajmują (najskromniej licząc) kilkadziesiąt gigabajtów, a więc **milion** razy więcej, niż opis tekstowy. Co więcej, pewnych informacji medycznych, które są widoczne i ewidentne na obrazach, nie powtarza się, ani nie komentuje w tekście, bowiem osoba sporządzająca dokumentację wychodzi z założenia, że jeśli dokumentację tę przeglądać będzie fachowiec (inny lekarz), to z obrazów sam wywnioskuje wszystko, co trzeba, więc nie warto o rzeczach oczywistych pisać w tekście, zwłaszcza że taki opis może czasem wprawić pacjenta w stan przerażenia – jak wiadomo, właśnie dla zachowania spokoju pacjenta lekarze unikają jawnego podawania w dokumentacji medycznej diagnoz i prognoz, posługując się w celu ich ukrycia między innymi terminologią łacińską. W dokumentacji elektronicznej ma to dodatkowe uzasadnienie związane między innymi z ryzykiem nieuprawnionego dostępu do danych ze strony ewentualnych hakerów.

a)

b)



Rys. 2. Rekord pacjenta w systemie TMIS – jednym z pierwszych systemów szpitalnych (a). Multimedialny rekord pacjenta na przykładzie systemu szpitalnego WorldVistA (b)
Źródło: a) [22]; b) [23]



Rys. 3. Obecnie często spotyka się sytuacje, w których główna treść przekazu informacyjnego zawiera się w obrazie, a nie w towarzyszącym mu tekście
Źródło: [24]

Sytuacja opisana wyżej w odniesieniu do danych medycznych powtarza się w przypadku różnych innych rodzajów danych, w których coraz więcej merytorycznie istotnych informacji przedstawianych jest wyłącznie (lub prawie wyłącznie) w formie obrazów (rys. 3).

Ze względu na wspomnianą wyżej informatywność obrazu, preferencje odbiorców informacji, a także coraz większą łatwość tworzenia cyfrowych obrazów – ten trend będzie się nasilał.

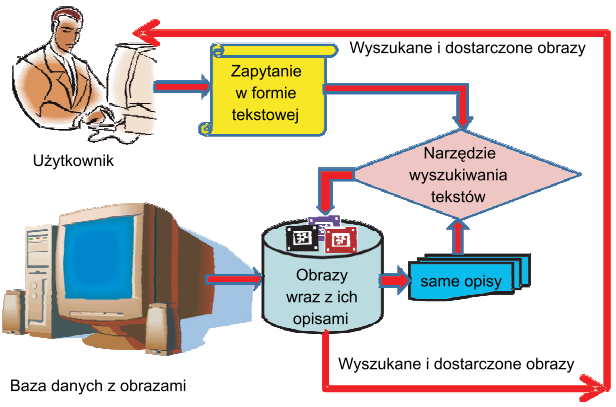
3. Problemy związane z używaniem informacji obrazowej

3.1. Wyszukiwanie potrzebnych informacji obrazowych

Obcowanie z pojedynczą informacją mającą postać obrazową jest w zasadzie wygodne i przyjemne. Mając na ekranie (lub na wydruku z komputera) ładnie zbudowany obrazek, możemy go przeanalizować (zakładając, że wiemy, co on przedstawia), następnie możemy na tej podstawie zdobyć pewne wiadomości, zaś przyswoiwszy i zapamiętawszy te wiadomości, możemy podjąć jakieś działania, które dzięki tej informacji będą skuteczniejsze.

Podany wyżej schemat wygodnego i przyjemnego korzystania z informacji obrazowej przestaje być aktualny z chwilą konfrontacji z bardzo dużą liczbą obrazów, wśród których trzeba znaleźć ten, który w danej chwili jest nam potrzebny. Wyszukanie potrzebnej informacji tekstowej w bazie danych zawierającej setki czy tysiące tekstów jest kłopotliwe, ale technicznie wykonalne. Korzystając ze słów kluczowych, możemy dotrzeć do potrzebnej informacji, biorąc za podstawę nawet tak obszerny zasób przeszukiwanych informacji jak cały Internet. Przykładem stosownego narzędzia, które tego rodzaju wyszukiwanie umożliwia, są popularne Google. Oczywiście jak wszyscy wiedzą przy wyszukiwaniu danych tekstowych napotyka się pewne trudności, na przykład związane z istnieniem synonimów oraz niejednoznacznością struktur języka naturalnego. Ale wiadomo także, iż trudności te można przezwyciężyć stosując między innymi bardzo dobrze rozwijającą się technikę ontologii i innych atrybutów tak zwanych sieci semantycznych [4]. Wiąże się to faktem, że wniknięcie do sfery znaczeniowej (semantyki) informacji tekstowej jest w miarę łatwe i raczej oczywiste [5].

Odmienna sytuacja ma miejsce, kiedy rozważamy duże zasoby informacyjne złożone z obrazów. W takim przypadku jeśli chcemy dotrzeć do jakiegoś konkretnego obrazu zawierającego jakąś konkretną treść – to możemy z tym mieć bardzo poważne kłopoty. Istniejące systemy wyszukiwania informacji oferują w zasadzie jedynie dwa typy narzędzi do wyszukiwania informacji obrazowej. Pierwsza z tych technik opiera się na opiach tekstowych, jakie posiadają obrazy w wielu miejscach sieciowych, skąd można je pobierać. Idealna sytuacja występuje w przypadku, gdy obraz ma podpis, a w tym podpisie występuje poszukiwane słowo (rys. 4).



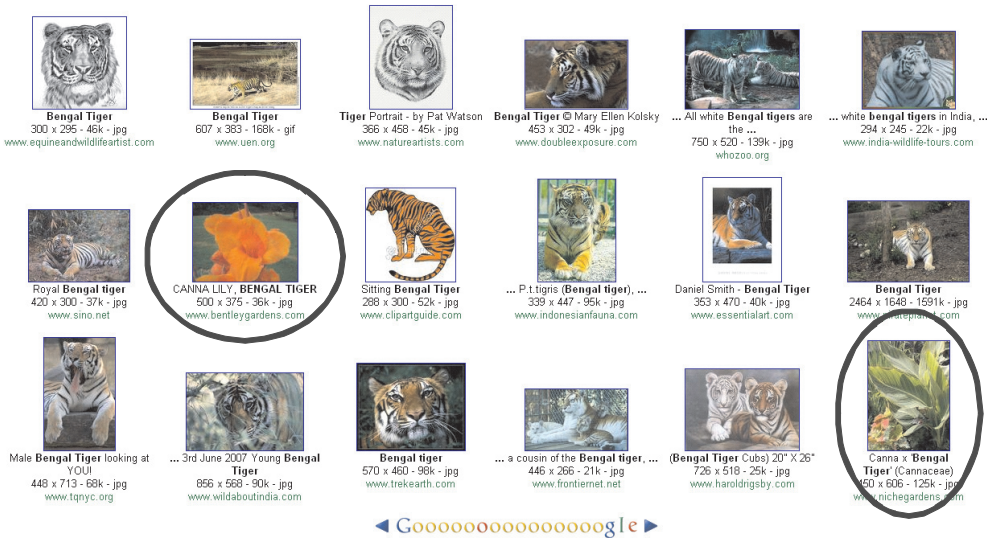
Rys. 4. Wyszukiwanie obrazów na podstawie ich podpisów

Systemy wyszukujące obrazy według podanego schematu podlegają wielu ograniczeniom, dlatego ich działanie jest z reguły dalekie od doskonałości. Po pierwsze – do czego jeszcze wrócimy – ich działanie ograniczone jest wyłącznie do takich baz z obrazami, w których obrazom towarzyszą opisy. Takie podejście można uznać za efektywne w odniesieniu do tych źródeł informacji, w których głównym nośnikiem treści jest tekst, a obrazów używa się wyłącznie jako ilustracji uzupełniających tę treść przekazywaną tekstem (w tym charakterze występują między innymi rysunki w niniejszym artykule). Jak jednak wspomniano wyżej, współcześnie pojawiają się w różnych bazach danych (a także w Internecie) zasoby informacyjne złożone głównie albo wyłącznie z obrazów, którym towarzyszą nieliczne i na ogół mało informatywne opisy tekstowe. W takim przypadku przedstawiona na rysunku 4 metoda wyszukiwania obrazów okaże się nieskuteczna.

Co więcej, łatwo można się przekonać, że nawet pozornie całkiem bezpieczne wyszukiwanie w dobrze opisanych zasobach informacji obrazowych na podstawie towarzyszących obrazom tekstów może prowadzić na manowce, czego przykładem może być wynik wyszukiwania przedstawiony na rysunku 5.

Hasłem wyszukiwanym była angielska nazwa *Bengal Tiger*. Oczywiście pokazały się rozliczne mordki tygrysów (i o to chodziło), ale nieoczekiwanie wśród znalezionych obrazów pojawiły się … kwiaty. Okazało się, że nazwę *Bengal Tiger* nosi także piękny kwiat rośliny zwanej po angielsku Canna (polska nazwa to kanna albo paciorecznik). Jak z tego widać, wyszukiwanie wykorzystujące związek tekstu i obrazu może czasem dawać dziwne skutki.

Systemy wyszukujące obrazy na podstawie ich tekstowych opisów mogą czasem zawieść z innych powodów, niż wcześniej wymienione. Wynika to z faktu, że programiści tych systemów starają się wywiązać ze stawianych im zadań w taki sposób, że nie ograniczają się do przeszukiwania samych tylko podpisów pod obrazami (bo w wielu zasobach obrazy jako takie nie są wcale podpisane), natomiast analizują tekst w otoczeniu obrazu, zakładając, że występowanie słowa kluczowego (użytego jako klucz wyszukiwania) w tekście, w którym obraz jest osadzony, wskazuje na to, że ten obraz i to słowo są ze sobą związane znaczeniowo.

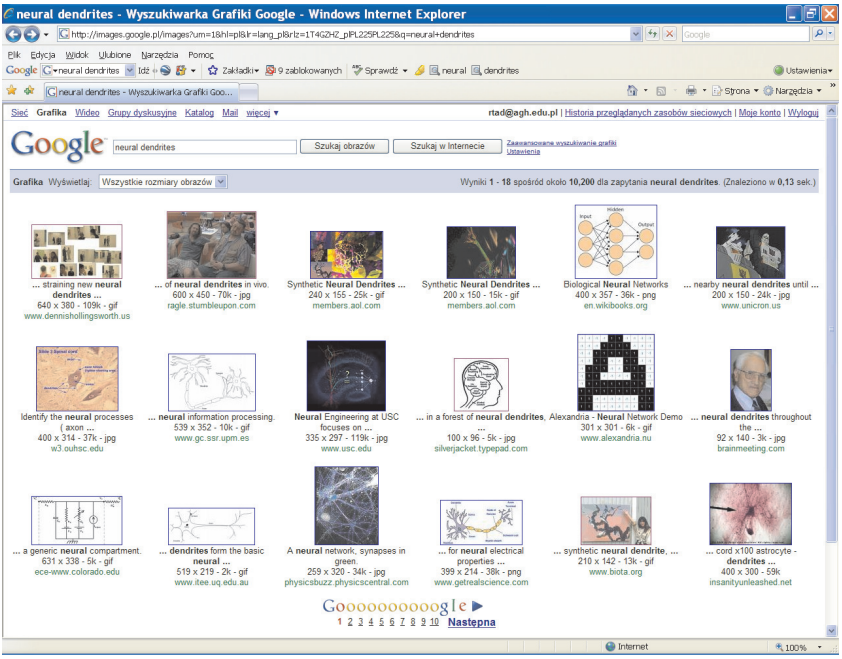


Rys. 5. Pomyłki przy wyszukiwaniu obrazów na podstawie opisów, zaobserwowane przy zastosowaniu wyszukiwarki Google

Niestety, założenie to nie zawsze sprawdza się w praktyce. Wyobraźmy sobie, że chcemy znaleźć (przykładowo) obraz dendrytów otaczających neuron. Wpisujemy w Google (w wersji wyszukującej obrazy) *neural dendrites* – i już po chwili mamy całą kolekcję obrazów (rys. 6) – w większości nie na temat!

Na czołowych miejscach listy znajdują się odwołania do zdjęć osób lub sytuacji, w których ktoś (pragnąc się wydać interesującym) użył określenia, że wysilał swoje dendryty neuronowe na przykład w tym celu, żeby prowadzić konwersacje na przyjęciu – a wyszukiwarka zaproponowała zdjęcie ludzi z tego przyjęcia jako zdjęcia dendrytów. Całkowity nonsens!

Można sformułować hipotezę, że przyczyną nieskutecznego wyszukiwania jest użycie niewłaściwego narzędzia. W końcu Google wyszukują strony WWW głównie na podstawie ich zawartości tekstowej, dlatego poszukiwanie adekwatnych merytorycznie obrazów może się zakończyć niepowodzeniem – i to niczego nie dowodzi. Spróbowaliśmy więc wyszukiwać obrazy z wykorzystaniem narzędzi opisywanych jako narzędzia specjalizowane do tego właśnie celu (to znaczy do wyszukiwania obrazów). Na rysunku 7 pokazano skutek wyszukiwania obrazów mających związek z medycyną z pomocą programu *PicSearch*, opisywanego jako właśnie specjalistyczne narzędzie do wyszukiwania obrazów. Widać, że zaraz na pierwszej stronie rezultatów dostarczonych przez wyszukiwarkę w zbiorze wyszukanych obrazów mamy przynajmniej jeden obraz całkowicie nie na temat. Podobny test z podobnym wynikiem można by było przeprowadzić dla każdego z dostępnych obecnie narzędzi wyszukiwawczych, gdyż wyszukiwanie **obrazów** na podstawie kryteriów zadanych w postaci **tekstu** po prostu nie może się udać.



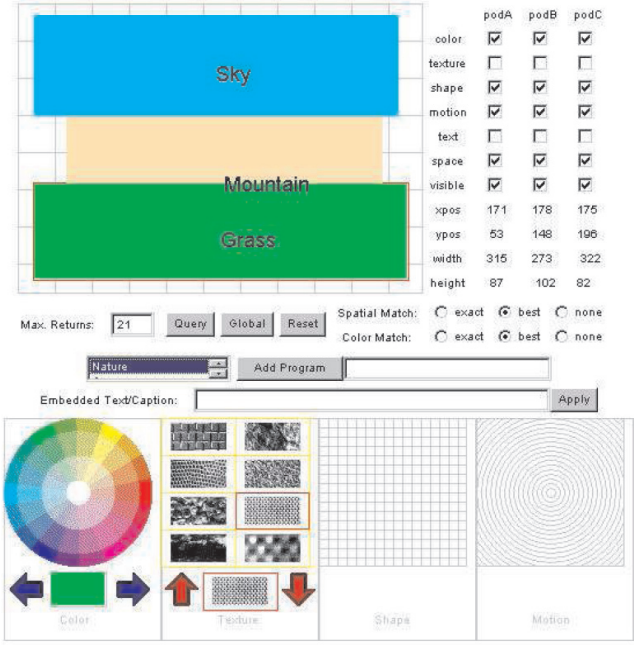
Rys. 6. Niepowodzenie przy wyszukiwaniu obrazów związanych z hasłem *neural dendrites*

Rys. 7. Nawet stosując specjalizowane narzędzia do wyszukiwania obrazów, bardzo często otrzymujemy nonsensowne wyniki

3.2. Próba wniknięcia w głąb obrazu

Opisane wyżej niepowodzenia, obserwowane wielokrotnie przy próbach odnajdywania potrzebnych obrazów na podstawie towarzyszących im tekstów, wskazują na to, że przy budowie przyszłych systemów informatycznych, zorientowanych na gromadzenie, przetwarzanie i wyszukiwanie informacji **obrazowej**, nie można będzie dalej traktować obrazów jako jedynie dodatku do informacji tekstowej, lecz trzeba będzie zmierzyć się z koniecznością wyszukiwania obrazów na podstawie cech tychże obrazów. Pojawia się tu jednak pytanie, jakie to powinny być cechy.

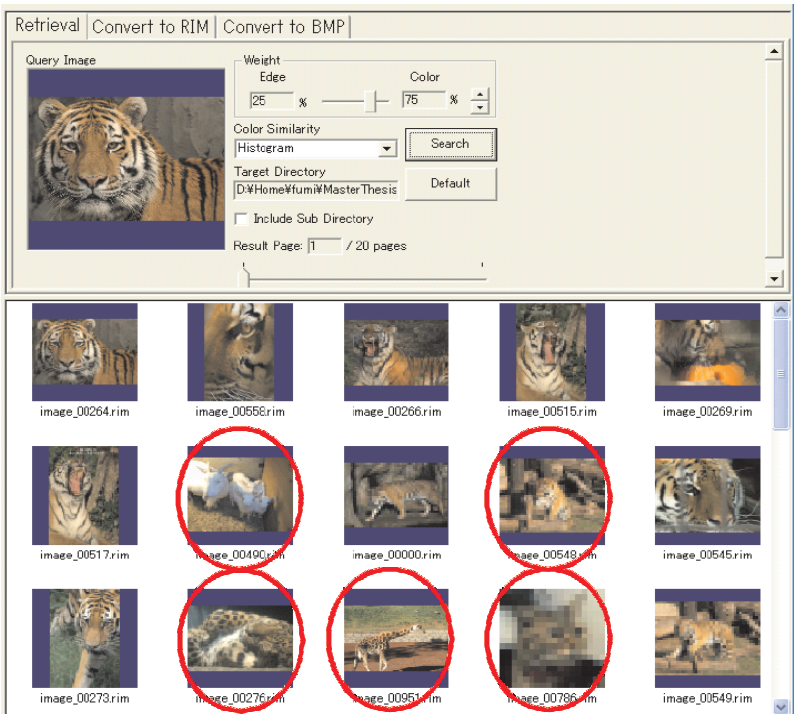
Biorąc pod uwagę różnorodność możliwych obrazów, znacznie większą niż różnorodność potencjalnie możliwych tekstów, trzeba zdać sobie sprawę z tego, że próba mechanicznego wyszukiwania obrazów na podstawie koincydencji jakichś wzorców graficznych jest raczej skazana na niepowodzenie. Przewidywanie to potwierdzają eksperymenty, jakie można przeprowadzić przy użyciu programu VisualSEEk [6]. Program ten pozwala podać ogólne kryteria na temat właściwości obrazów, które mają być wyszukane, w formie przedstawionej przykładowo na rysunku 8, a następnie pozwala wyszukiwać w bazie danych wszystkie obrazy, które spełniają wskazane kryteria. Warto zauważyć, że kryteria, jakie są podane (przykładowo) na rysunku 8, są kryteriami ściśle odnoszącymi się do właściwości obrazu – podaje się wzorce kolorów, jakie powinny dominować w określonych regionach obrazu, przykładowe tekstury, cechy kształtu obiektów itp.



Rys. 8. Przykład zadawania kryteriów wyszukiwania obrazów na podstawie cech graficznych
Źródło: [6]

Niestety wyniki nadal nie są zadowalające. Okazuje się, że w dużych bazach danych mogą występować liczne obrazy spełniające podane wyżej kryteria, a jednak nie odpowiadające wymaganiom osoby formułującej zapytanie (która, jak widać, zapewne chciała wydobyć górskie pejzaże).

Równie iluzoryczna jest nadzieja na to, że problem wyszukania obrazów na jakiś założony temat uda się rozwiązać poprzez zaprezentowanie systemowi wyszukującemu przykładowego obrazu, jaki chcemy wyszukać. Na rysunku 9 pokazano zapytanie realizowane metodą *query by example* (zapytaniem jest obraz przedstawiający tygrysa) w systemie Wisvi [7] oraz skutki wyszukiwania zainicjowanego tym zapytaniem. Widać, że komputerowa ocena podobieństwa obrazów, będąca podstawą do ich wyszukania i udostępnienia, nie całkiem zgadzała się w rozważanym przypadku z oceną, jakiej dokonałby człowiek na podstawie analizy zawartości przykładowego obrazu opartej na swojej **wiedzy** o tym, co przedstawia przykład i co zawierają obrazy przeszukiwanej bazy danych. Jest to kolejny argument przemawiający za tym, że przy operowaniu obrazami nie wystarczy skupiać się na ich formie, lecz trzeba koniecznie sięgnąć do treści, usiłować wydobyć i **zrozumieć** ich znaczenie.

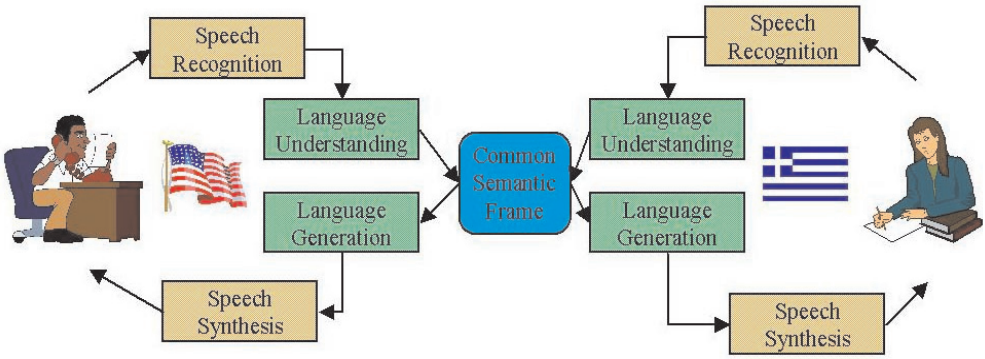


Rys. 9. Nie całkiem udana próba wyszukiwania obrazów
na podstawie ich podobieństwa do wzorca podanego w zapytaniu

3.3. Co w istocie oznacza rozumienie obrazu?

Postulat, jakim zakończyliśmy poprzedni podrozdział, wskazuje, że przy inteligentnym wyszukiwaniu danych w bazach danych, w których istota informacji zawarta jest w postaci obrazów – trzeba dążyć do tego, żeby komputer docierał do **znaczenia** obrazu, czyli zmierzał do jego **automatycznego rozumienia**.

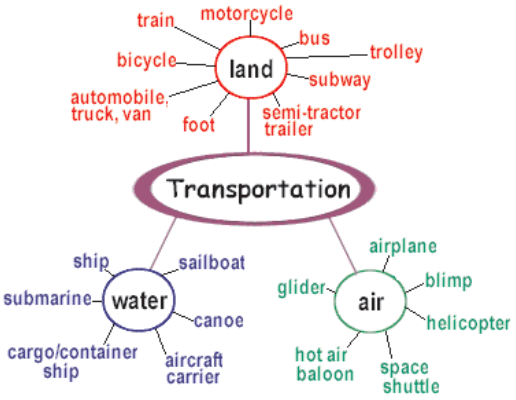
Pojęcia automatycznego rozumienia nie jest w informatyce czymś szokująco nowym, bowiem liczne prace związane z pogłębioną analizą tekstów (w kontekście między innymi tak zwanych sieci semantycznych [2, 4]) wprowadziły to pojęcie do obiegu i nadały mu określone znaczenie, wykorzystywane na przykład w systemach automatycznego tłumaczenia tekstów z jednego języka na drugi (rys. 10).



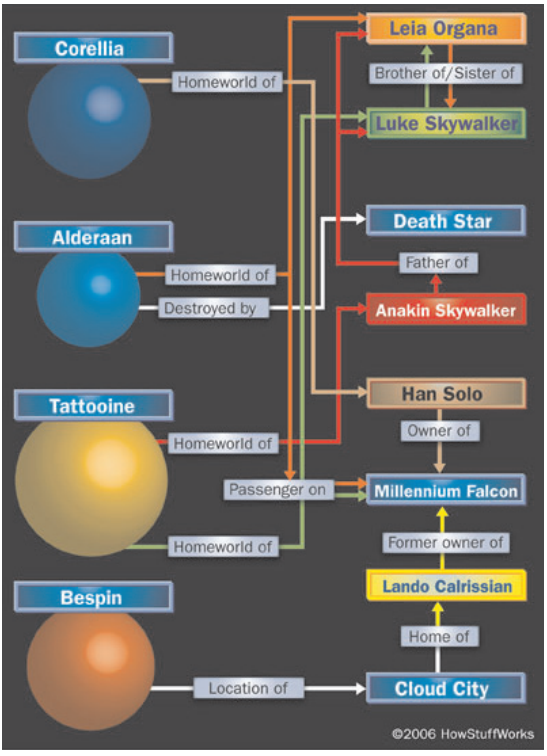
Rys. 10. Rozumienie wypowiedzi jest kluczem do automatycznego tłumaczenia z jednego języka na inny bez utraty zrozumiałości treści
Źródło: [25]

Narzędzia do automatycznego rozumienia tekstów są dziś dosyć mocno rozbudowane i są wykorzystywane w różnych celach, na przykład do automatycznego wyszukiwania tekstów mających związek z jakimś zagadnieniem (niemożliwych do wyszukania za pomocą słów kluczowych), automatycznego streszczania dokumentów, tworzenia reprezentacji wiedzy itd. Podstawą do ich działania są elementy specjalnie reprezentowanych w komputerze merytorycznych związków pomiędzy informacjami, zadawanych za pomocą tak zwanych ontologii. Przykład takiej bardzo prostej ontologii przedstawia rysunek 11.

Inny przykład nieco bardziej złożonej ontologii (sporządzonej na potrzeby miłośników „Gwiezdnych wojen") obejrzeć można na rysunku 12.



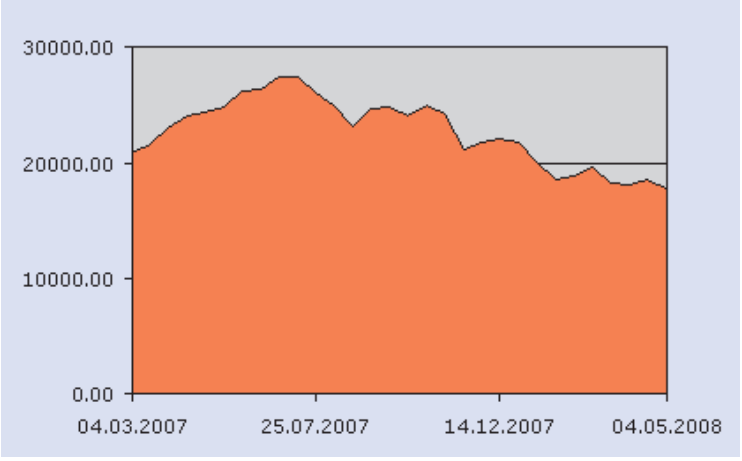
Rys. 11. Przykład bardzo prostej ontologii przydatnej przy automatycznym rozumieniu tekstów
Źródło: [26]



Rys. 12. Przykładowy obraz semantycznych zależności występujących w pewnej grupie wyrazów
Źródło: [27]

Jednak rozumienie (a nawet **automatyczne** rozumienie) tekstów jest relatywnie łatwe, naturalne i bezkonfliktowe, ponieważ każdy spodziewa się w nich zawartości określonej treści, dzięki czemu rozważanie ich w kategorii ich **znaczeń** nikogo nie szokuje. Natomiast wyraźnie czymś innym jest rozumienie obrazów. Obraz ma określoną **formę**: rozmiar, kolor, fakturę, kształty obiektów itd. W tych kategoriach obraz można opisać i na przykład w sposób automatyczny rozpoznać – jeśli zajdzie taka potrzeba. Ale na pierwszy rzut oka wydaje się, że obraz nie ma czegoś takiego jak **treść**, która jest w znacznym stopniu niezależna od formy. Czy zatem można mówić o **rozumieniu obrazów**?

Zanim odpowiemy na to pytanie – posłużymy się kilkoma przykładami. Rozważmy na początku rysunek 13.

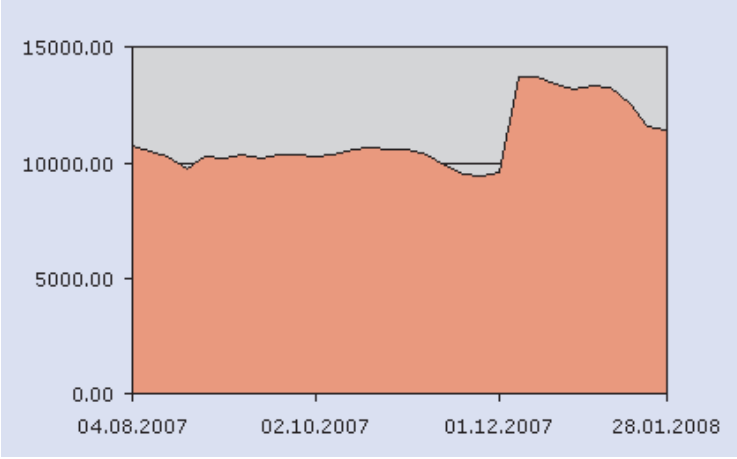


Rys. 13. Obraz, który może być łatwo zinterpretowany w sensie jego znaczenia

Na rysunku tym pokazany jest jakiś wykres, którego znaczenia na pozór nie można zrozumieć, jeśli nie ma do niego dodatkowego opisu. Jeśli jednak przyjrzymy się temu rysunkowi uważniej, to zauważymy, że opis osi poziomej przypomina daty, a opis osi pionowej kojarzy się z pieniędzmi. **Łącząc** te dwie **obserwacje** z **wiedzą**, jaką prawie każdy z Czytelników posiada na temat „zawirowań" na giełdzie papierów wartościowych na przełomie 207 i 2008 roku, można **zrozumieć,** że oglądamy spadek wartości jakiejś inwestycji giełdowej. Gdyby chcieć to zrozumienie podsumować jednym krótkim słowem, to można by było napisać: ***bessa*.**

Na rysunku 14 widoczny jest bardzo podobny wykres, który jednak może być podstawą bardziej wyrafinowanego **rozumienia** pewnego zdarzenia. Otóż na wykresie tym (który na podstawie analogii z wcześniej rozważanym obrazem identyfikujemy także jako obraz zmian wartości jakiegoś waloru notowanego na giełdzie) da się zaobserwować ciekawsze zjawisko. Widać na nim, jak po długiej serii spadków notowania zaczęły nieznacznie wzrastać (w okolicy daty 6 grudnia, czyli Świętego Mikołaja).

Inwestor zachęcony tym faktem przewidywał, że zaczął się trend wzrostowy i dokupił za kilka tysięcy walorów, co spowodowało gwałtowny skok wykresu do góry. Niestety nadzieja na wzrost (czy chociażby stabilizację kursu) okazała się zwodnicza – akcje zamiast wzrastać zaczęły gwałtownie spadać i cała dopłata została w istocie stracona. Gdyby chcieć to zrozumienie znaczenia pokazanego wykresu podsumować jednym krótkim słowem, to można by było napisać: ***błąd inwestycyjny***.



Rys. 14. Obraz, który może być podstawą zrozumienia pewnego zdarzenia

Warto zauważyć, że na to lakonicznie spuentowane zrozumienie sytuacji złożyły się dwie składowe: to, co można było **zobaczyć** na obrazie (wykresie) oraz to, co obserwator **wiedział** o naturze procesów zachodzących na giełdzie papierów wartościowych. Dopiero konfrontacja tych dwóch źródeł informacji dała podstawę do **zrozumienia** treści, jakie niósł rozważany obraz. Jednocześnie fakt, że na podstawie samego obrazu (oraz posiadanej wiedzy) można było dotrzeć do tej merytorycznej treści – wskazuje na to, że obraz może zawierać określoną treść i że można się starać tę treść **zrozumieć.** A skoro człowiek może zrozumieć tę treść – to jest możliwe, że także komputery potrafią to zrobić, albowiem w ciągu wielu lat rozwoju sztucznej inteligencji (obecnie nazywanej inteligencją obliczeniową) wielokrotnie wykazano, że ilekroć jakąś formę intelektualnej aktywności człowieka udało się dobrze zdefiniować i precyzyjnie opisać na gruncie psychologii, a zwłaszcza kognitywistyki, tylekroć po krótkim czasie udawało się także zbudować program komputerowy, który tę aktywność intelektualną człowieka potrafił naśladować – a w wielu wypadkach nawet wyprzedzać.

Dyskusja obrazków pokazanych na rysunkach 13 i 14 dowiodła, że w obrazach może być zwarta określona treść, a odpowiednia analiza obrazu, połączona z wykorzystaniem określonych zasobów wiedzy – może prowadzić do tego, że treść ta zostaje wydobyta i **zrozumiana.** Jednak sceptyczny krytyk, wątpiący w możliwość **rozumienia obrazów**

oraz docierania do ich sensu – może nie być przekonany podanymi przykładami z tego powodu, że obrazy na rysunkach 13 i 14 były sztucznymi wykresami, rejestrującymi jakieś zdarzenia, natomiast mało mającymi wspólnego z obrazami pochodzącymi z realnego świata. W takich sztucznych obrazach ludzie (znający się na rzeczy) z łatwością doszukują się określonych znaczeń, ale to niczego nie dowodzi, bo te obrazy tak właśnie były tworzone, żeby te ustalone znaczenia eksponować. W kolejnym podrozdziale zajmiemy się obrazami, które pochodzą z realnego świata, a jednak ponad wszelką wątpliwość niosą w sobie znaczenia, których nie można po prostu **zobaczyć** na obrazie, tylko trzeba je właśnie **zrozumieć.**

3.4. Jak rozumieć obrazy realnego świata?

Zacznijmy znowu od przykładu. Popatrzmy na fotografię przytoczoną na rysunku 15.

Rys. 15. Obraz, który wymaga **zrozumienia,** a nie tylko prostego przeanalizowania
Źródło: [28]

Na fotografii tej widać znane przedmioty – skrzynki z napojami oraz butelki – całe oraz potłuczone. Tło dla tych obiektów też jest znane – po prostu droga wiodąca przez las. A jednak mimo tych znanych obiektów i znanej scenerii coś jest w tym obrazie niepokojącego. Po prostu te obiekty w tej scenerii w taki sposób nie powinny występować! Mówi

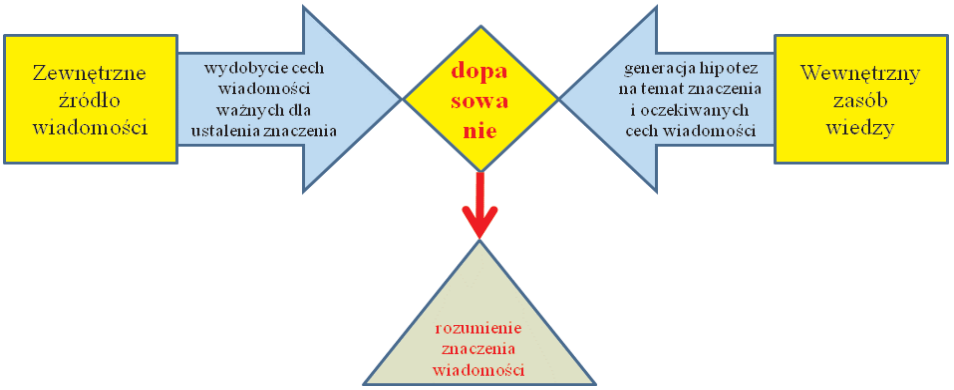
nam to **wiedza**, jaką posiadamy na temat tego, w jakich okolicznościach i gdzie można widywać butelki piwa oraz ich skrzynki, a także na temat tego, jak powinna wyglądać normalna leśna droga. Dlatego wnikając w **znaczenie** tego obrazu (a nie poprzestając na samej jego formie), możemy **zrozumieć**, co ten obraz naprawdę oznacza – mianowicie możemy stwierdzić, że te skrzynki musiały spaść z jadącej ciężarówki. Na zdjęciu nie widać tej ciężarówki, a jednak wiemy na pewno, że musi ona tam być, podobnie jak nie byliśmy świadkami samego zdarzenia, ale potrafimy sobie wyobrazić jego przebieg. Mało tego – możemy odpowiedzieć na pytanie, co było prawdopodobną przyczyną tego wypadku i kto ponosi odpowiedzialność.

Przykładów, w których obraz zawiera (poza formą) jakąś merytoryczną treść, można by jeszcze podać bardzo wiele. Nie chodzi jednak o mnożenie przykładów tylko o rozwiązanie problemu, **w jaki sposób komputer może dotrzeć do tej sfery znaczenia obrazu, ignorując sferę jego formy?**

Odpowiedzią na tak sformułowane pytanie jest technika automatycznego rozumienia obrazów, którą w Katedrze Automatyki AGH rozwijamy już od około 10 lat, a która w skrócie przedstawiona jest niżej.

4. Koncepcja automatycznego rozumienia obrazów

Ważnym elementem każdego **rozumienia** jest okoliczność, że dla wniknięcia w semantyczną treść określonej wiadomości – trzeba **zawartość** wiadomości jako takiej skonfrontować z **wiedzą** posiadaną przez rozumiejący podmiot. Dotyczy to wiadomości w każdej postaci (także w postaci obrazu) oraz każdego podmiotu, który ma zmierzać do osiągnięcia stanu określanego jako rozumienie, nawet jeśli podmiotem tym jest komputer, a wynikiem jego działania ma to być analiza kognitywna określana jako automatyczne rozumienie.



Rys. 16. Ogólny schemat automatycznego rozumienia

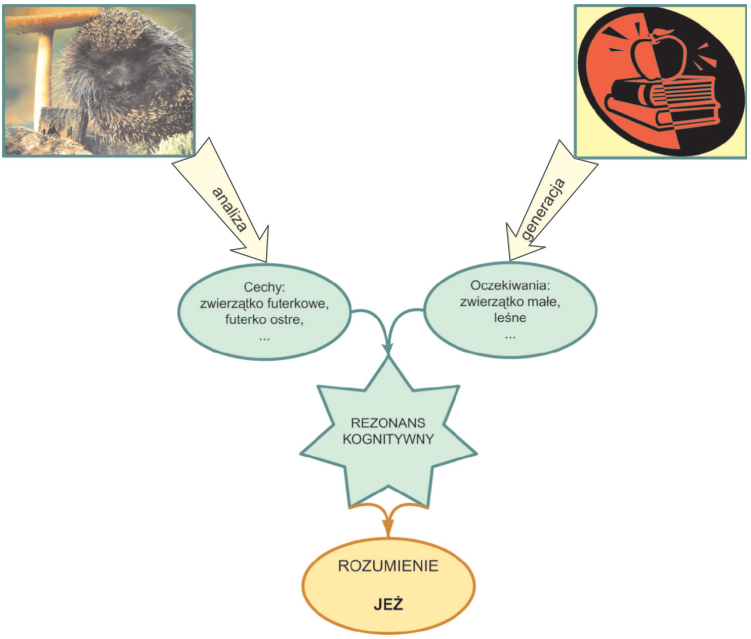
Najbardziej ogólny schemat rozumienia może więc być rozpatrywany w takiej formie, jaka została przedstawiona na rysunku 16. Schemat ten może być użyty dla rozumienia wiadomości dowolnego rodzaju.

W odniesieniu do obrazów stosunkowo niewielu autorów (poza autorami niniejszej pracy, o czym będzie dodatkowo mowa nieco dalej) próbowało podjąć problem ich automatycznego rozumienia. Warto wspomnieć o pracy [8] dotyczącej jednak rozumienia obrazów bardzo specjalnego rodzaju (zapisu nutowego) oraz o pracy [9], reprezentującej jednak podejście analityczno-geometryczne do próby zrozumienia kształtów, a nie takie, jakie proponuje się w tej pracy, preferujące strukturalny (lingwistyczny) opis całego obrazu. Ciekawe wyniki zawiera praca [10], skupiająca się jednak wyłącznie na obrazach jednego rodzaju (tęczówki oka dla celów biometrycznej identyfikacji osób). Również o próbach rozumienia obrazów osób (głównie w sensie ich identyfikacji) traktuje praca [11], aczkolwiek prezentowane w niej podejście jest całkowicie inne od tego, jakie dyskutujemy w tym artykule. Natomiast podejście zbliżone do prezentowanego tutaj (to znaczy oparte na konfrontacji nadchodzącej informacji wizyjnej z wiedzą wcześniej zgromadzoną w systemie) znaleźć można w pracach [12, 13].

5. Opis lingwistyczny jako baza automatycznego rozumienia obrazów

Po dokonaniu tego skrótowego przeglądu literatury spróbujemy przedstawić istotę proponowanej przez autorów koncepcji automatycznego rozumienia obrazów. Wyjdziemy od schematu podanego na rysunku 16 i spróbujemy go w kilku kolejnych krokach adaptować do specyficznych potrzeb, jakie rodzi zadanie automatyczne rozumienie obrazów. We wcześniejszych pracach autorów oraz Pani dr Lidii Ogieli (patrz m.in. [14]) zaproponowano schemat wymyślony przez tę ostatnią, nieco bardziej ukierunkowany na cel, jakim jest automatyczne rozumienie właśnie obrazów (a nie na przykład wiadomości tekstowych). Schemat ten przedstawiamy na rysunku 17. Warto zwrócić uwagę na kilka elementów tego schematu, które chcemy tu skomentować.

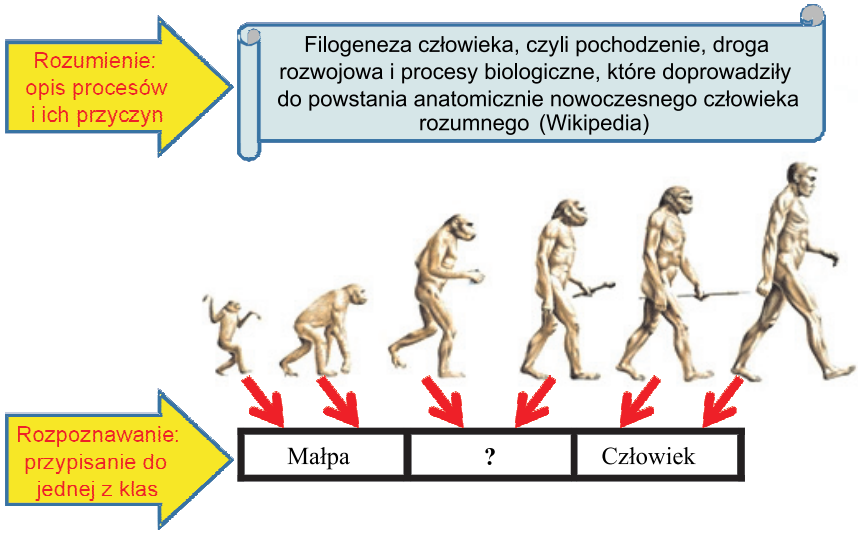
Po pierwsze w stosunku do rysunku 16 pojawiło się tu nieco więcej konkretów w zakresie zarówno kanału wejściowego (pochodzącego od obrazu pozyskanego za pomocą kamery), jak i wewnętrznego łącza informacyjnego, przenoszącego oczekiwania generowane przez zdeponowany w systemie zasób wiedzy. Jak widać, wejściowy obraz powinien być w sposób specjalny reprezentowany w systemie, który ma go **zrozumieć**, a nie tylko mechanicznie przetworzyć albo rutynowo przeanalizować (z ewentualnym przejściem do automatycznego rozpoznawania, które jednak jest czymś innym, niż automatyczne rozumienie). Obraz podlegający komputerowej obróbce, której celem ma być zrozumienie jego zawartości, musi być w systemie prezentowany z wykorzystaniem metod lingwistycznych, czyli musi być przedstawiony jako łańcuch symboli terminalnych pewnej gramatyki.



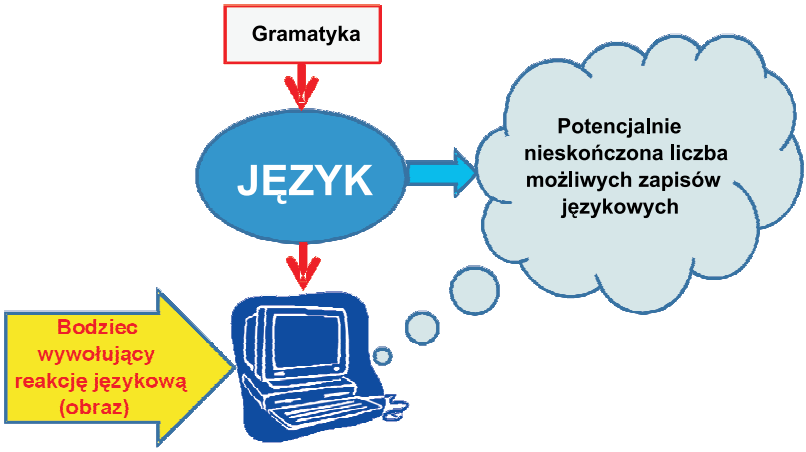
Rys. 17. Sposób postępowania z obrazami zmierzający do automatycznego rozumienia ich zawartości

Warto skupić się na chwilę nad pytaniem, dlaczego do opisu obrazów przeznaczonych do analizy, której celem jest automatyczne rozumienie używane są narzędzia lingwistyczne. Otóż powód jest związany z samą naturą procesu rozumienia, którą warto w tym miejscu odróżnić od natury procesu rozpoznawania, z którym automatyczne rozumienie jest dosyć często mylone. Patrząc na rysunek 18, widzimy ogólnie znaną sekwencję istot żywych. Gdy naszym zadaniem jest rozpoznawanie, wówczas najpierw ustalamy listę możliwych klas, do których można zaliczyć analizowane obiekty. Lista taka zawsze ma skończoną, z góry określoną liczbę pozycji (wliczając w to zazwyczaj pozycję „nie wiadomo", oznaczoną na rysunku znakiem zapytania), zaś zadaniem algorytmu analizującego obraz jest stwierdzenie, do której z tych wcześniej przewidzianych klas należy zaliczyć ten czy inny konkretny obiekt.

Natomiast **rozumienie** obrazu (osiągane przez inteligentnego człowieka, studiującego obraz, lub uzyskiwane automatycznie, do czego zmierzają badania referowane w tej pracy) oznacza wydobycie z obrazu tych wszystkich znaczeń, które są w nim *implicite* zawarte, ale nie są *explicite* widoczne (patrz rys. 18 w jego górnej części). Warto zauważyć, że w odróżnieniu od rozpoznawania, dla którego zbiór odpowiedzi systemu jest z góry zdeterminowany, w przypadku rozumienia sposób interpretacji obrazu jest nieprzewidywalny i z tego powodu zbiór możliwych opisów obrazu jest potencjalnie nieskończony. Jest to poważna trudność, gdyż tę potencjalnie nieskończoną różnorodność musi wytworzyć narzędzie o bezspornie skończonych możliwościach – komputer (rys. 19).



Rys. 18. Różnica między rozpoznawaniem i rozumieniem obrazu



Rys. 19. Język jako narzędzie umożliwiające wytworzenie potencjalnie nieskończonej liczby zapisów przez urządzenie o skończonej liczbie możliwości (komputer)

Otóż język jest właśnie takim narzędziem, które pozwala na generowanie nieskończenie różnorodnych kombinacji, bazujących na skończonej liczbie elementów. Na przykład język polski składa się ze skończonej liczby słów i oparty jest na gramatyce mającej skończoną liczbę reguł – a jednak można w nim napisać nieskończoną liczbę artykułów, powieści, poematów, pism urzędowych itp. Również języki sztuczne (na przykład **C++**) cechują się tym, że mając skończoną liczbę składników oraz reguł (łatwą do opanowania przez

komputerowy kompilator) – mogą służyć do wytworzenia nieograniczonej liczby programów, potencjalnie nieskończonej, po napisaniu dowolnej liczby programów zawsze możliwe jest napisanie jeszcze jednego, kolejnego.

6. Podsumowanie

Ograniczone rozmiary tej pracy nie pozwalają na wniknięcie we wszystkie szczegóły sygnalizowanej tu koncepcji automatycznego rozumienia obrazów. Gramatyki służące do opisu obrazów podlegających potem automatycznemu rozumieniu opisane zostały między innymi w pracach [15, 16], a następnie było to obszernie przedyskutowane w książce [17].

Pojęcie rezonansu kognitywnego zostało wprowadzone w pracy [18], a obszernie przedyskutowane było w książce [19]. Zainteresowanego czytelnika odsyłamy do tych właśnie pozycji literatury, albo zachęcamy do poszukiwania najnowszych prac z tego zakresu poprzez wpisanie w Google hasła *Automatic understanding of the images.*

Mimo wielu prac, o których wspomnieliśmy w treści artykułu, technika automatycznego rozumienia obrazów znajduje się na bardzo wstępnym etapie swego rozwoju. Poza opisanymi dokładniej w cytowanej literaturze (patrz także prace [20, 21]) zastosowaniami w medycynie, które mają jednak swoją specyfikę – metodologia automatycznego rozumienia obrazów w praktyce jeszcze nie istnieje. Opisane w tej pracy ogólne zasady mogą stanowić punkt wyjścia do opracowania takiej metodyki, a także mogą stanowić zachętę do podejmowania dalszych prac na ten temat.

Literatura

[1] Kaptelinin V., Czerwinski M. (eds.), *Beyond the desktop metaphor*. Cambridge MA, The MIT Press 2007.
[2] Sieckenius de Suza C., *The semiotic engineering of human-computer interactions.* Cambridge MA, The MIT Press 2005.
[3] Stiny G., *Shape*. Cambridge MA, The MIT Press 2008.
[4] Antoniou G., van Harmelen F., *A semantic web primer.* Cambridge MA, The MIT Press 2008.
[5] Tadeusiewicz R., *Nowoczesne metody kognitywnej analizy danych ekonomicznych i dokumentów tekstowych oraz ich zastosowanie w zarządzaniu przedsiębiorstwem*. Rozdział w pracy zbiorowej: Waszkielewicz W.: Zarządzanie organizacjami w gospodarce rynkowej, Kraków, UWND 2007, 239–260.
[6] Smith J.R., Shih-Fu Chang, *VisualSEEk: a fully automated content-based image query system*. 1996, Dostępny http://www.ee.columbia.edu/ln/dvmm/researchProjects/ MultimediaIndexing/VisualSEEk/acmmm96/acmfin.html.
[7] Usage of Wisvi 2007, http://www.wiz.cs.waseda.ac.jp/~rim/usage-e.html.
[8] Homenda W., *Automatic understanding of images: integrated syntactic and semantic analysis of music notation*. Proceedings of International Joint Conference on Neural Networks, Vancouver, 2006, 3026–3033.
[9] Leś Z., Tadeusiewicz R., *Shape Understanding System, Polygon Class Processing Methods*. In Hamza M.H. (ed.): "Signal Processing and Communications", Anaheim, Calgary, Zurich, IASTED/ACTA Press 2000, 447–454.

[10] Bowyer K.W., Hollingsworth K., Flynn P.J., *Image understanding for iris biometrics: A survey.* Computer Vision and Image Understanding, Vol. 110, Issue 2, 2008, 281–307.
[11] Hilton A., Fua P., Ronfard R., *Modeling people: Vision-based understanding of a person's shape, appearance, movement, and behaviour*. Computer Vision and Image Understanding, vol. 104, Issues 2-3, 2006, 87–89.
[12] Drummond T., Caelli T., *Learning Task-Specific Object Recognition and Scene Understanding*. Computer Vision and Image Understanding, vol. 80, Issue 3, 2000, 315–348.
[13] Jiang Yu Zheng, Saburo Tsuji, *Generating Dynamic Projection Images for Scene Representation and Understanding*. Computer Vision and Image Understanding, vol. 72, Issue 3, 1998, 237–256.
[14] Ogiela L., Tadeusiewicz R., Ogiela M.R., *Cognitive techniques in medical information systems.* Computers in Biology and Medicine, vol. 38, Nr 4., 2008, 501–507.
[15] Ogiela M.R., Tadeusiewicz R., *Image Understanding Methods in Biomedical Informatics and Digital Imaging*. Journal of Biomedical Informatics, Computers and Biomedical Research, vol. 34, No. 6, 2001, 377–386.
[16] Tadeusiewicz R., Ogiela M.R., *Automatic Understanding of Medical Images – New Achievements in Syntactic Analysis of Selected Medical Images*. Biocybernetics and Biomedical Engineering, vol. 22, nr 4, 2002, 17–29.
[17] Ogiela M.R., Tadeusiewicz R., *Modern Computational Intelligence Methods for the Interpretation of Medical Image*. Studies in Computational Intelligence, vol. 84, 2008, Springer-Verlag, Berlin – Heidelberg – New York.
[18] Ogiela M.R., Tadeusiewicz R., *Artificial Intelligence Structural Imaging Techniques in Visual Pattern Analysis and Medical Data Understanding*. Pattern Recognition, vol. 36/10, 2003, 2441–2452.
[19] Tadeusiewicz R., Ogiela M.R., *Medical Image Understanding Technology*. Series: Studies in Fuzziness and Soft Computing, vol. 156, 2004, Springer-Verlag, Berlin – Heidelberg – New York.
[20] Ogiela M.R., Tadeusiewicz R., Ogiela L., *Image Languages in Intelligent Radiological Palm Diagnostics*. Pattern Recognition, vol. 39/11, 2006, 2157–2165.
[21] Ogiela M.R., Tadeusiewicz R., Trzupek M., *Graph-based semantic description and information extraction in analysis of 3D coronary vessels visualizations*. In: Badica C., Paprzycki M. (eds.): Advances in Intelligent and Distributed Computing, Studies in Computational Intelligence, vol. 78, 2008, Springer-Verlag, Berlin – Heidelberg – New York, 303–309.
[22] Policy Implications of Medical Information Systems, Office of Technology Assessment, Congress of the United States, 1977
[23] WorldVistA Monograph http://www.va.gov/vista_monograph
[24] http://www.forex.nawigator.biz/wykresy-punktowo-symboliczne.php
[25] http://oxygen.csail.mit.edu/images/Speech.jpg
[26] http://www.kidbibs.com/images/semantic.gif
[27] http://static.howstuffworks.com/gif/semantic-web-4.jpg
[28] http://www.maryannhorton.com/images/horrible-accident.jpg