

ANNA OLSZEWSKA  
JOANNA GANCARCZYK

## WYSZUKIWANIE WIZUALNE W OBIEGU KULTURY: STRATEGIE PROJEKTOWE, PRZYKŁAD SYSTEMU PRINTEXPLORER

Niniejszy komunikat prezentuje koncepcję oraz podstawowe zasady działania systemu umożliwiającego rozpoznawanie obrazów. Odwołujemy się do doświadczeń wyniesionych z prac nad projektem PrintExplorer. System wyszukiwania wizualnego został zaprojektowany z myślą o kolekcji znajdującej się w zbiorach Biblioteki Naukowej PAU i PAN w Krakowie. Jak wiele podobnych sobie systemów eksperckich, służy do rozpoznawania specyficznego rodzaju form, takich jak znaki własnościowe, monogramy czy sygnatury widoczne na cyfrowych kopiach historycznych grafik, starodruków i map.

Szczegóły dotyczące rozwiązań technicznych projektu przedstawione zostały we wcześniejszych publikacjach jego auterek. Celem niniejszego sprawozdania jest pokazanie samej strategii projektowania tego rodzaju narzędzi. Nakreślenie skali zmian, które będziemy w stanie wprowadzać, gdy porzucając na moment domenę technologii, oprzemy strategię projektową na tezach współczesnej humanistyki, odwołamy się do studiów nad kulturą. Podstawą prac stanie się wówczas obserwacja praktyk komunikowania wizualnego oraz uwzględnienie procesów długiego trwania w odniesieniu do określonych sposobów organizacji wiedzy.

### ZAŁOŻENIE I. PROJEKTOWANIE KONTEKSTUALNE: BUDOWA SYSTEMÓW UMOŻLIWIĄJĄCYCH SWOBODNE ŁĄCZENIE ZASOBÓW

Obserwacja procesów długiego trwania, a w szczególności archeologia wiedzy, dowodzi, że współczesne systemy informacji nie są jednorodne. Nowe kolekcje muzeów powstają zazwyczaj ze zlepiania starszych zbiorów, poprzez donacje, w drodze decyzji politycznych, zmian historii. Również katalogi tych kolekcji po-

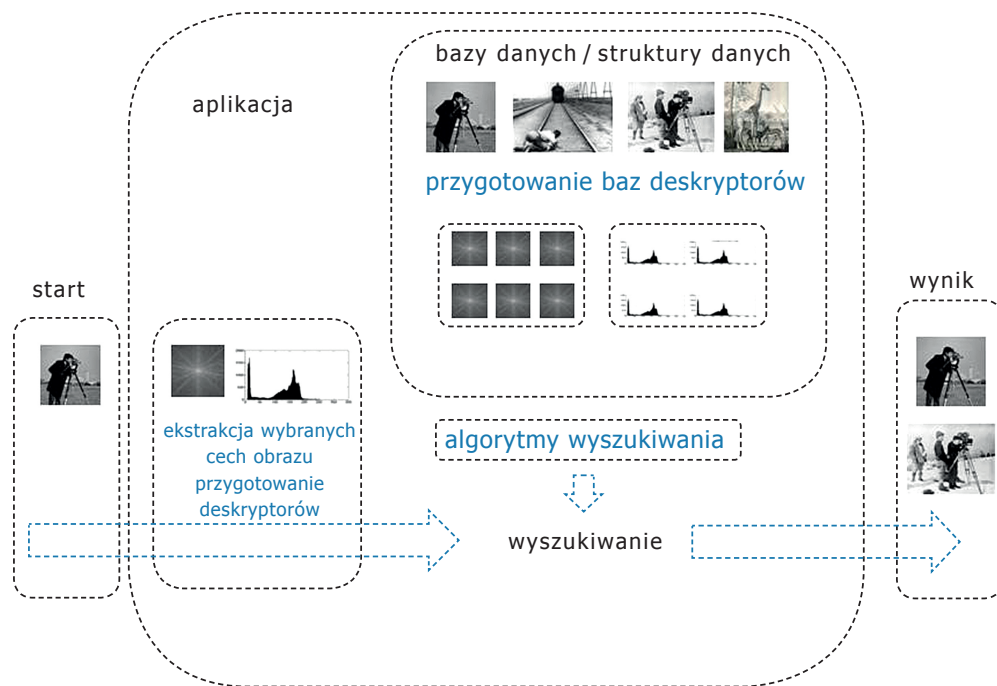
zostają nieciągłe, funkcjonują w kilku wersjach równocześnie, stale przepisywane, nigdy nie realizują zunifikowanego wzorca. Ta naturalna różnorodność dotyczy również domeny cyfrowej. Digitalizując zbiory, biblioteka tworzy nową zamkniętą całość, opierającą się zazwyczaj na działaniu osobnej bazy danych. W tych warunkach system utrzymywany jest na ogół w sposób zachowawczy, działa dzięki systematycznemu porządkowaniu jego struktury. Bibliotekarze niechętnie negocjują jego zmiany. Migracje baz postrzegane są jako operacje wymagające specjalnej mobilizacji. Wprowadzenie nowych elementów (np. wyszukiwania wizualnego) do tak zorganizowanego obiegu zazwyczaj wymaga ingerencji w jego strukturę, jest pracochłonne, oceniane jako ryzykowne. Z tego powodu projektowanie nowych narzędzi humanistyki cyfrowej, jeżeli ma do siebie przekonać przyszłych użytkowników, powinno zakładać konieczność nieinwazyjnego funkcjonowania w zastanym systemie cyfrowych kolekcji.

W pracach nad projektem PrintExplorer uwzględniono obowiązujący w bibliotece podział kompetencji oraz wypracowany w ramach jej działania system digitalizacji i katalogowania zbioru. Mając na uwadze konieczność łączenia zasobów kultury w jeden obieg, założono projektowanie narzędzia autonomicznego. System będzie współdziałał z cyfrową reprezentacją kolekcji PAU w sposób bezszwowy. W konsekwencji umożliwi to dodawanie do wyszukiwania kolejnych repozytoriów tego rodzaju. Pomysł wymaga odejścia od klasycznego schematu bazy danych. Sposób na zredukowanie jej znaczenia odnaleziono w obserwacji praktyk związanych z katalogowaniem i opisywaniem zabytków muzealnych.

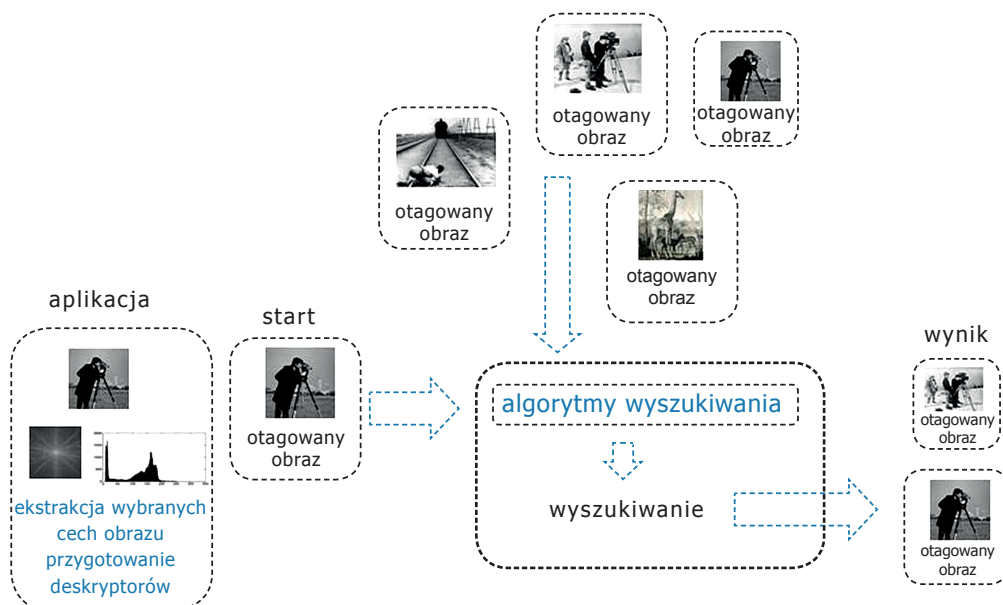
Założono, iż informacje potrzebne do porównania obrazów przechowywane będą nie przez administratora bazy danych, ale zostaną zapisane bezpośrednio w metadanych pliku wizualnego<sup>1</sup>. Operacja, którą nazwać możemy tagowaniem form wizualnych, podobna jest wcześniejszym praktykom katalogowania. Obraz zamieszczony w cyfrowym repozytorium opisać możemy na wiele sposobów. Do stworzenia takiego opisu używamy szeregu słów: tematycznych haseł (tagów), odniesień do techniki, miejsca wykonania, autorstwa itd. Niekiedy dodajemy do niego zapisy numeryczne, umożliwiające np. geolokację dokumentu. Obrazowi takiemu przyporządkować możemy również określone deskryptory wizualne. Będą to reprezentacje kształtów, barw, tekstur, zapisane w formie ciągów numerycznych. Dzięki ich odnotowaniu będziemy w stanie wyszukiwać obrazem. Ponieważ tagi wizualne przypisane będą do publikowanego pliku, opierając się na mechanizmie stosowanym w obiektowych bazach danych i języku XML, nie będzie konieczne budowanie osobnej bazy danych, złożonej z pomierzonych obrazów. Ciągi cyfr wskazujące na dany obraz będzie w stanie wyłowić w sieci zwykła wyszukiwarka tekstowa.

---

<sup>1</sup> A. Olszewska, J. Gancarczyk, *Wyszukiwanie wizualne w obiegu kultury: deskryptory i tagowanie* [w:] *Badanie i projektowanie komunikacji*, seria „projektowanie komunikacji”, t. 6, Wrocław 2017, s. 177–191.



Ryc. 1. Klasyczny schemat systemu CBIR (porównywanie obrazów z przykładami zamieszczonymi w bazie danych)



Ryc. 2. Schemat autonomiczny, zakładający tagowanie form wizualnych

## ZAŁOŻENIE II. WYKORZYSTANIE WIELOZNACZNOŚCI OBRAZÓW: INTERFEJSY I INGERENCJA UŻYTKOWNIKA

Większość współcześnie używanych systemów wyszukiwania wizualnego pozostawia użytkownikom niewielkie możliwości definiowania istotnych w danym momencie fragmentów obrazu. Interakcja zazwyczaj ograniczona jest do możliwości kadrowania danego przykładu. Metoda utrudnia niekiedy dokładniejsze zdefiniowanie poszukiwanego kształtu. Jest też czasochłonna. Rozpowszechnienie ekranów dotykowych wykształciło w użytkownikach nowe nawyki. Przyzwyczajono nas do nawigowania palcami, ciągłego dotyknięcia powierzchni obrazu, tekstu, odnośników. Obserwacja nowych praktyk otwiera możliwość przeprojektowania interfejsów wyszukiwania graficznego. Zaznaczając fragment obrazu za pomocą kilku prostych ruchów, użytkownik zyskuje większą możliwość kontroli procesu wyszukiwania. Tak skonstruowany interfejs PrintExplorera daje *de facto* możliwość zadawania pytań w porządku wizualnym<sup>2</sup>. Oddawanie kontroli użytkownikom jest strategią przeciwną do tendencji zakładającej całkowitą automatyzację wyszukiwania wizualnego. Pozostaje jednak w zgodzie z dążeniem do dookreślenia wieloznacznej natury obrazu, w zależności od danego kontekstu użycia.

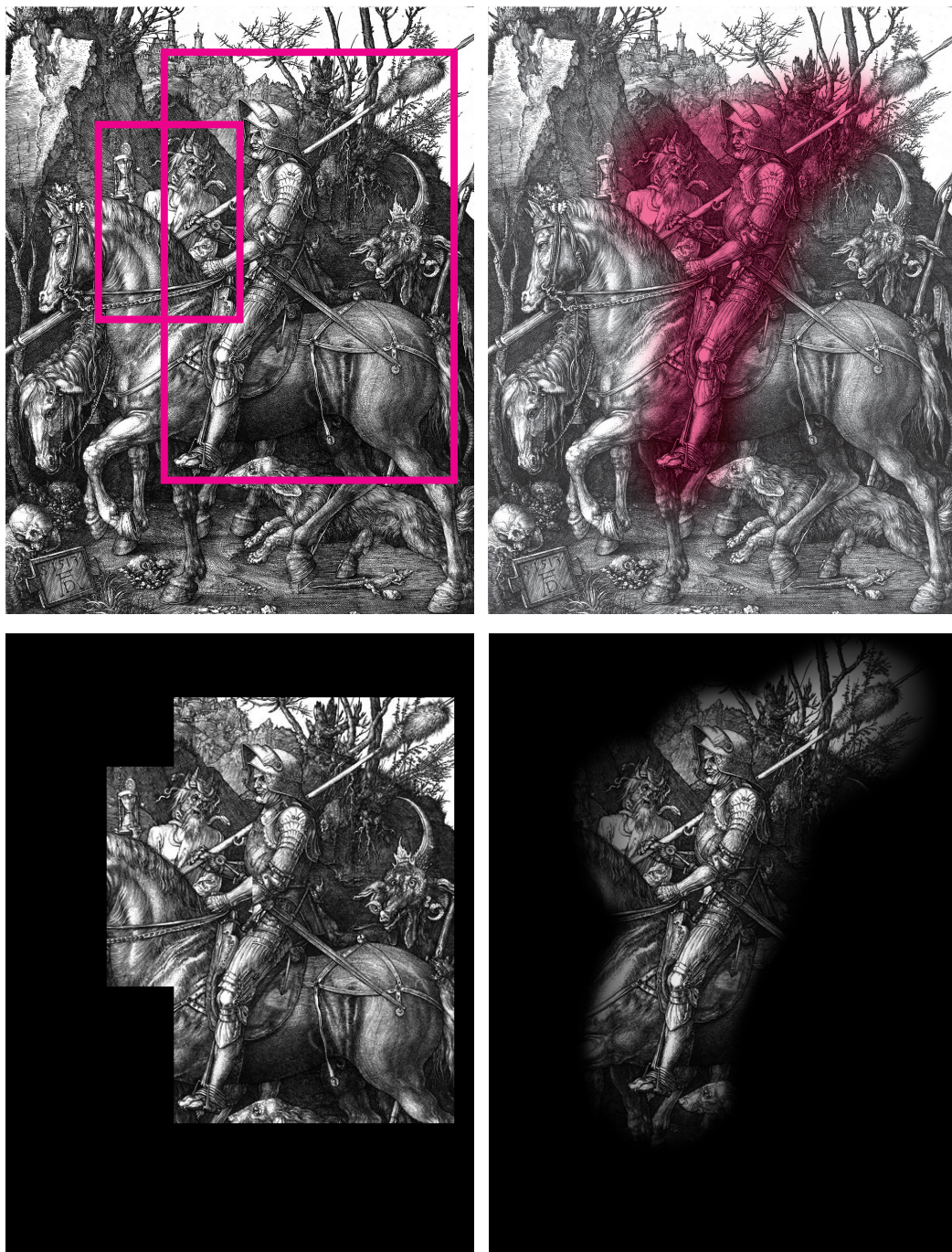
Strategia obserwacji praktyk komunikowania wizualnego może być również wykorzystana przy tworzeniu interfejsów wyjściowych<sup>3</sup>. Klasycznym rozwiązaniem jest prezentowanie wyników wyszukiwania w porządku liniowym. Po zadaniu pytania otrzymujemy listę, uporządkowaną według kluczy zaprojektowanych przez twórców serwisu. Ponieważ zapisane alfabetem słowa same w sobie są gotowymi kawałkami kodu, określenie zasad tworzenia takiego rankingu w sytuacji wyszukiwania semantycznego wydaje się dużo bardziej przejrzyste. Może znaleźć swoje podstawy w definiowaniu i statystyce wybranych słów kluczy. Operując obrazami, wkraczamy w obszar trudny do zdefiniowania. Każdy element komunikatu wizualnego znaczyć może różnie, każdy algorytm uwypuklić może różne wymiary tego samego przykładu. Różnice zaczynają się już w chwili, gdy decydujemy o wyszukiwaniu wedle kształtu, tekstury czy gamy barwnej. W tej sytuacji projektanci serwisu, nie chcąc narzucać jednego z góry zadanego klucza użytkownikom, stoją przed zadaniem takiego przeorganizowania sposobów prezentowania wyników, które pozwoli na odnalezienie pożądaných obrazów bez konieczności zagłębiania się z zawilości zasad działania procedur algorytmicznych. W rozwiązaniu problemu wykorzystać można znajomość podstawowych konwencji budowania map i infografik. Doświad-

---

<sup>2</sup> A. Olszewska, J. Gancarczyk, *Touchscreen User Interface Design for Content Based Image Retrieval*, „Proceedings of EVA London 2017, UK”. Published by BCS Learning and Development Ltd. DOI: <http://dx.doi.org/10.14236/ewic/EVA2017.64>. 315.

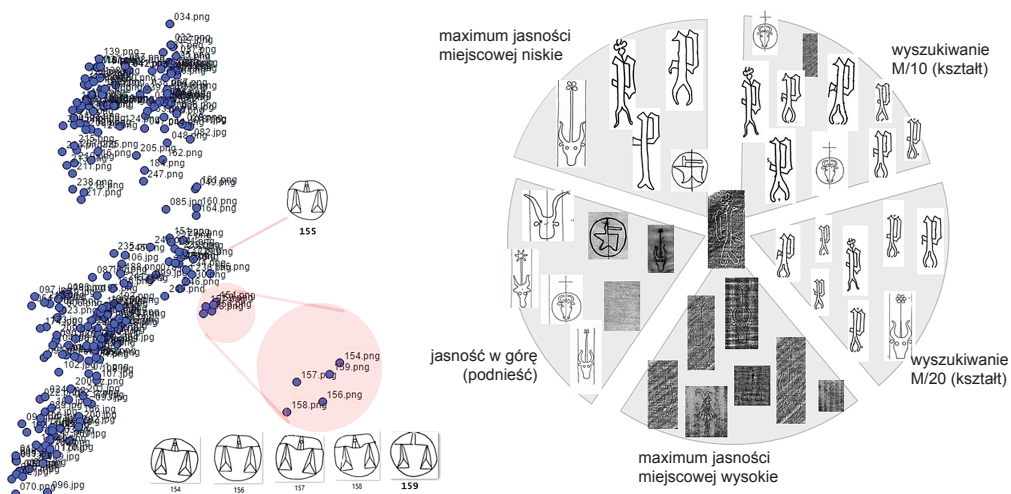
<sup>3</sup> J. Gancarczyk, A. Olszewska, *Which Entry is More Similar? A Non-linear Visualisation of Query Results in Image Retrieval and Image Recognition Problem*, „Electronic Visualisation and the Arts, EVA 2017, London, UK, July 11–13, 2017”, BCS 2017, Workshops in Computing, <https://doi.org/10.14236/ewic/EVA2017.14>.





Ryc. 3. Wskazanie ważnych fragmentów obrazu: kadrowanie vs zaznaczenie swobodne

czenie badań nad kulturą wizualną podpowiada, że tego rodzaju formy są łatwe do zaadaptowania. Prezentacja wyników wyszukiwania w formie klastrów, stosowanie układów centralnych pozwalają na uruchomienie wyuczonych schematów czytania obrazu (np. od centrum ku granicom, blisko = podobne). Wychodząc poza schemat linearny, właściwy tekstom, zaprezentować można wyniki porównania nie jednego, ale kilku kluczy wyszukiwania jednocześnie, pozostawiając decyzję o wyborze właściwej metody samemu odbiorcy.



Ryc. 4. Przykłady prezentacji wyników: wizualizacja t-SNE; układ centralny z wynikami pięciu rankingów

\* \* \*

Zasygnalizowane wyżej rozwiązania oparto na interpretacjach praktyk i konwencji funkcjonujących w przestrzeni kulturowej. Zaproponowane zmiany dotyczą sposobu budowania interfejsów, organizacji systemu przechowywania wyników. Nie dotyczą samych metod obliczania podobieństwa obrazów. Niemniej jednak, mają wpływ na ich wybór oraz efektywność ich działania. Wprowadzenie możliwości swobodnego oznaczania wybranych fragmentów ułatwi proces segmentacji obrazów (oddzielanie rysunku od tła, definiowanie granic form). Równoległe prezentowanie wyników stosowania kilku metod obliczeniowych może zostać sprzężone z trenin- giem sieci neuronowej rozpoznającej obrazy (jako informację zwrotną dla syste- mu wykorzystać można reakcje użytkowników zaznaczających pożądane wyniki). Proponowana zmiana strategii projektowych wykracza poza zakres definiowany przez standardowe badanie doświadczeń użytkowników (User Experience, front- end design).

Proponowane przeformułowanie metod projektowania niesie za sobą również spore ryzyko. Oddanie kontroli użytkownikowi oraz próba budowy systemu dzia-

łającego bez jasno zdefiniowanej bazy danych nie dają szansy na realizację z góry założonych standardów. Nie mamy pewności, jakie elementy obrazu zaznaczać będą użytkownicy. Niemal na pewno niemożliwe będzie idealne powtórzenie takiego zestawu gestów. Nie mamy gwarancji co do tego, że tagowanie form wizualnych obejmie całość danego zbioru. Nie jesteśmy w stanie przewidzieć wyborów, jakich dokonają użytkownicy przeglądający równoległe prezentowane wyniki rankingów, co za tym idzie, nie wiemy, jak będzie przebiegał trening sieci regulującej pracę systemu. Jesteśmy zmuszeni działać na rozmytych zbiorach danych. Podporządkowując się różnorodności doświadczeń wizualnych zakładamy jednakże, że z tych działań wyłoni się pewna regularność. Studia semiologiczne dawno temu ogłosiły fiasko poszukiwań alfabetu wizualnego. Jednak sama możliwość utrzymywania tego rodzaju komunikacji dowodzi istnienia pewnych zjawisk regulujących świat obrazów. Projektowany system, podobnie jak i współczesne studia wizualne, odchodzi od wizualnego esencjonalizmu, zakładając pewną stałość praktyk komunikowania wizualnego.