

Witold Abramowicz, Tomasz Kaczmarek, Marek Kowalkiewicz

Akademia Ekonomiczna w Poznaniu

SEKWENCJONOWANIE OBIEKTÓW UCZĄCYCH W INTELIGENTNYCH SYSTEMACH NAUCZANIA NA ODLEGŁOŚĆ

1. Wstęp

Sekwencjonowanie szkoleń jest technologią analizowaną już od dłuższego czasu w dziedzinie inteligentnych systemów uczących. Stoi za nim idea tworzenia zindywidualizowanych ścieżek szkoleń dla studenta poprzez adaptacyjny dobór najbardziej relewantnych oraz optymalnych pod innymi względami obiektów szkoleniowych w dowolnym etapie edukacji. Obiekt szkoleniowy jest definiowany jako „dowolny zasób cyfrowy, który może być użyty do wspomagania procesu nauczania” [6]. Należy rozróżnić dwa poziomy sekwencjonowania:

- sekwencjonowanie ścieżek szkoleniowych, polegające na budowaniu całej ścieżki szkoleniowej (nauczanie z celem długookresowym) z wykorzystaniem większych obiektów uczących (całe szkolenia i moduły),
- sekwencjonowanie pojedynczych szkoleń, polegająca na tworzeniu jednego szkolenia z wykorzystaniem drobnoziarnistych, atomowych, obiektów szkoleniowych.

Sekwencjonowanie jest jedną z najbardziej popularnych technologii w rozwijanych inteligentnych systemach uczących, jednakże komercyjne systemy zarządzania szkoleniami (CMS – Course Management System) wyjątkowo rzadko wykorzystują tę technologię (potwierdza się to w przypadku tak popularnych systemów, jak TopClass oraz WebCT). Najczęściej komercyjne systemy zarządzania szkoleniami pozwalają jedynie na statyczne sekwencjonowanie – poprzez dostarczanie jedynie statycznych, nieadaptowalnych sekwencji obiektów szkoleniowych.

1.1. Definicja poprzednika

Sekwencja obiektów szkoleniowych, która ma zostać wykorzystana w nauczaniu, jest tworzona na podstawie pewnej specyfikacji zależności pomiędzy obiektami szkoleniowymi. Są one najczęściej nazywane poprzednikami (szkoleniami poprzedzającymi). Najprostsza definicja poprzednika mogłaby brzmieć: **jeżeli pojęcie a jest wymagane do zrozumienia pojęcia b , to pojęcie a jest poprzednikiem pojęcia b** lub bardziej ogólnie, **jeżeli zawartość (obiektu uczącego) a jest wymagana do zrozumienia zawartości (obiektu uczącego) b , to a jest poprzednikiem b** . Inna definicja poprzednika, bardziej odpowiednia w przypadku adaptowalnych systemów hipermedialnych (AHS – Adaptive Hypermedia System), jest następująca: **jeżeli element x jest poprzednikiem elementu y , to x powinien być odwiedzony przed y** . Można w związku z tym wyciągnąć wniosek, iż **jeżeli a jest wymagane do zrozumienia b , to a powinno być odwiedzone przed b** . Jednakże takie założenie może zaprzeczać specyfikacjom w modelu pedagogicznym adaptowalnego systemu hipertekstowego [2]. W pewnych podejściach pedagogicznych, takich jak np. *problem based learning* lub *learning by doing*, oczekuje się, że studenci samodzielnie rozpoznają swoje braki w wiedzy, a przez to lepiej rozumieją swoje potrzeby. W rezultacie R. Huebscher rozróżnia dwa typy poprzedników: mechanizmy porządkujące (lub warunki wstępne) oraz poprzedniki pedagogiczne. W proponowanej metodzie zapisu wykorzystujemy poprzedniki jako mechanizm porządkujący, ale mimo to uważamy, że może ona być przydatna także do reprezentowania poprzedników pedagogicznych.

1.2. Przykładowa formalna reprezentacja zależności między szkoleniami

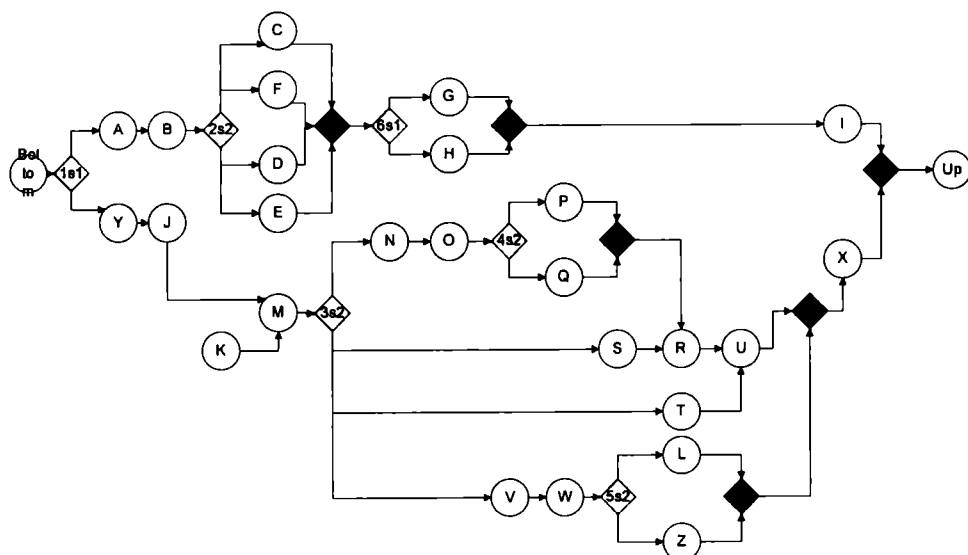
Obecny poziom rozwoju technologii służących do reprezentacji wiedzy (m.in. sieci semantyczne, ontologie) pozwala na wykorzystanie jednego z dostępnych języków reprezentacji do formalnego przedstawienia sekwencji szkoleń, dlatego przykładową formalną reprezentację należy traktować jedynie jako sugestię wykorzystania metod semantycznych. Do celów formalnej reprezentacji sekwencji szkoleń wykorzystujemy acykliczny graf skierowany, zawierający trzy typy węzłów: węzły standardowe, węzły początkowe oraz węzły końcowe (rys. 2).

Rysunek 1 przedstawia formalną reprezentację zależności pomiędzy szkoleniami. Nawiasy kwadratowe zostały zamienione na węzły dwóch rodzajów:

- węzły początkowe (s), w których podana została liczba obiektów szkoleniowych wymaganych do nauczania (oznaczone literą s; nazwa węzła składa się z numeru węzła, informacji o typie węzła i liczby wymaganych obiektów szkoleniowych);
- węzły końcowe (e), w których sprawdzane jest spełnienie warunków podanych w odpowiadających im węzłach początkowych.

Ze względu na dziedzinę zastosowania w tego typu grafach jest stosunkowo mało połączeń. Jest to zaleta, gdyż złożoność obliczeniowa algorytmów wykorzystujących sekwencje obiektów szkoleniowych jest znaczna. Zakładamy również

brak cykli w grafie reprezentującym ścieżki szkoleń. Jeżeli użytkownik zapoznał się z danym obiektem szkoleniowym, nie potrzebuje poznawać go ponownie. Oczywiście inna sytuacja zachodzi, gdy użytkownik testuje swoją wiedzę lub zapomina materiał. W takim wypadku konieczne okazać się cykle. Spójność grafu zapewniona jest poprzez wprowadzenie wierzchołka *tabula rasa*, oznaczającego zerowy stan wiedzy. Wszystkie węzły w grafie są nazwane, węzły startowe niosą ze sobą dodatkową informację. Istotną cechą grafu jest to, że każdy węzeł startowy może mieć tylko jednego poprzednika, a każdy węzeł końcowy – tylko jednego następnika. Jeżeli potrzebne są dodatkowe węzły, oznacza to, iż konieczne jest dodatkowe zagnieżdżanie.



Rys. 1. Przykładowa formalna reprezentacja sekwencji szkoleń

Źródło: opracowanie własne.

2. Metoda automatycznego tworzenia sekwencji szkoleń oraz sekwencji pojęć

Poprzedni rozdział prezentuje podejście do zapisywania zależności między obiektami uczącymi oraz opisywania ich sekwencji. Niniejszy rozdział stanowi prezentację oryginalnej metody, która może wspomagać sekwencjonowanie obiektów uczących oraz dostarczyć informacji ułatwiającej nawigowanie między szkoleniami oraz pojęciami w inteligentnym systemie nauczania.

2.1. Problem automatyzacji sekwencjonowania

W chwili obecnej sekwencje obiektów uczących tworzą metodycy na podstawie swojej znajomości tych dokumentów oraz dziedziny wiedzy. W dobie rosnącej

liczby obiektów uczących dostępnych w środowisku internetu oraz stron internetowych mogących stanowić uzupełniający materiał dydaktyczny rodzi się **problem automatycznego sekwencjonowania obiektów uczących**.

Ze względu na złożoność problemu oraz niedoskonałość obecnych metod semantycznego indeksowania oraz wykorzystania ontologii wydaje się nierealne, aby sekwencjonować obiekty uczące w pełni automatycznie. Sekwencjonowanie szkoleń może być jednak wspomagane przez wyznaczenie zależności między dokumentami, będących w relacji poprzednik – następnik.

Na zależności między dokumentami wpływają wykorzystane w dokumentach pojęcia (łatwiejsze lub trudniejsze z punktu widzenia uczącego się) oraz inne składniki, takie jak użyte sformułowania, zastosowana metoda dydaktyczna itp. Niezwykle trudno jest określić wpływ tych wszystkich składników na ocenę, czy dany dokument jest „łatwiejszy” czy też „trudniejszy”, i w związku z tym na ich potencjalną sekwencję. Do dalszych rozważań wybrano więc wyłącznie analizę występowania pojęć w dokumentach jako dającą się przeprowadzić np. przez analogię do metod wyszukiwania informacji w tekstach (*information retrieval*).

Proponowana metoda wyznaczania zależności między dokumentami ma na celu wspomaganie sekwencjonowania oraz nawigacji między obiektami uczącymi. W dalszej części tekstu posługujemy się uproszczoną relacją zależności między dokumentami, opierając się jedynie na częstości występowania w nich pojęć. Wygodną reprezentacją takiej relacji może być skierowany ważony graf acykliczny, w którym dokumenty są wierzchołkami, natomiast zwrot oraz waga na krawędziach są interpretowane jako siła relacji między dokumentami.

2.2. Zależności między pojęciami

Sekwencja takich pojęć odzwierciedla przekonanie metodyka o kolejności, w jakiej powinien być studiowany dany przedmiot. Warto tu zwrócić uwagę, że metodyki mogą się różnić między sobą, tzn. uznawać za poprawne różne sekwencje poznawania pojęć. Wśród metodyk nauczania znajdują się zarówno takie, które wymagają, aby przed użyciem nowego pojęcia wprowadzić jego definicję (używając pojęć wcześniej znanych), jak i takie, które zakładają zdefiniowanie pojęcia dopiero po jego wykorzystaniu, najczęściej w przykładach ilustrujących jego znaczenie. Istnienie różnych podejść umotywowanych dydaktycznie nie wpływa jednak na sam fakt istnienia relacji pomiędzy pojęciami polegającej na definiowaniu pojęć bardziej złożonych z wykorzystaniem prostszych lub należących do sfery wiedzy powszechnej.

2.3. Relacja obiektów uczących a relacja pojęć

Między relacją poprzedzania zdefiniowaną na materiałach dydaktycznych a relacją poprzedzania zdefiniowaną na pojęciach zachodzi silna analogia. Jest ona

związana z tym, że każdy materiał dydaktyczny wykorzystuje lub dotyczy pewnych pojęć, które można odnieść do jego treści (poszczególnych zwrotów i wyrazów w niej występujących). Relacje te (relacja poprzedzania na materiałach dydaktycznych oraz relacja poprzedzania na pojęciach) można przetransformować jedną do drugiej lub odwrotnie. W praktyce zazwyczaj dysponujemy przynajmniej jedną z wymienionych relacji:

- dane są sekwencje zdefiniowane na szkoleniach (w przypadku wnętrza pojedynczego obiektu uczącego lub w przypadku gotowej sekwencji szkoleń – kursu),
- dostępna jest relacja zdefiniowana na pojęciach (określona np. przez metodyków przedmiotu).

Możliwość transformacji jednej z relacji w drugą można wykorzystać w systemach informacyjnych wspierających dydaktykę na kilka sposobów:

- mając dane zależności między szkoleniami można w sposób automatyczny lub półautomatyczny stworzyć ich katalog (wykorzystując metody analizy skupień lub kategoryzacji) oraz wskazać w nim powiązania między pojęciami, które ułatwiają uczniom nawigowanie w materiale dydaktycznym,
- mając daną relację na pojęciach oraz kolekcję nieuporządkowanych materiałów potencjalnie przydatnych w dydaktyce, można zasugerować sekwencje tych materiałów zgodne z daną metodyką.

Takie sekwencje mogą być następnie wzbogacane o wyfiltrowane z zewnętrznych źródeł (w tym ze źródeł internetowych) dodatkowe materiały, które można automatycznie umieścić na właściwej pozycji w sekwencji, wzbogacając w ten sposób cały materiał dydaktyczny o najnowsze treści. Mając relację określoną na pojęciach można też uzupełniać braki w materiałach dydaktycznych – o ile takie występują – zwracając się do specjalistów o przygotowanie materiałów dotyczących zadanych pojęć.

2.4. Wyznaczanie zależności między obiektami uczącymi

Aby uzyskać relację określoną na zbiorze obiektów uczących, musimy dysponować relacją określoną na zbiorze pojęć. Dla uproszczenia przyjmujemy następujące założenia, nieograniczające jednak ogólności dalszego wywodu:

- obiekty uczące są obiektami tekstowymi,
- pojęcia są termami występującymi w tych dokumentach lub można je łatwo do takich termów odnieść,
- w przedstawionej metodzie opieramy się na modelu wektorowym systemu wyszukiwawczego, w którym kluczowa dla reprezentacji dokumentu jest częstotliwość występowania termów w nim zawartych.

Można zatem łatwo ocenić, w jakim stopniu każde pojęcie dotyczy każdego z dokumentów, obliczając liczebności pojęć i wyznaczając ich częstotliwości występowania w dokumentach (np. według jednego z przyjętych w literaturze dotyczącej systemów wyszukiwawczych schematu ważenia termów).

W niniejszej metodzie posługujemy się częstotliwością wystąpienia termu w dokumencie:

Częstotliwości występowania pojęć (termów) w obiektach uczących tworzą macierz F , której wiersze odpowiadają kolejnym dokumentom (1, ..., n), a kolumny – pojęciom (1, ..., m)

$$F = \begin{Bmatrix} f_{11} & f_{1m} \\ \dots & \dots \\ f_{n1} & f_{nm} \end{Bmatrix}.$$

Elementy macierzy F oblicza się, korzystając ze wzoru:

$$f_{ij} = \frac{c_{ij}}{d_i}.$$

Gdzie: f_{ij} – częstotliwość występowania termu j w dokumencie i ,

c_{ij} – liczba wystąpień termu j w dokumencie i ,

d_i – liczba wystąpień wszystkich termów w dokumencie i (długość dokumentu).

Relację zależności między pojęciami można przedstawić za pomocą macierzy W , której elementy odpowiadają sile zależności między pojęciami. Element w_{ij} to waga zależności między pojęciem i a pojęciem j .

Jako maksymalną siłę relacji między pojęciami przyjmujemy 1,0 (lub -1,0 zależnie od kierunku relacji), jednakże można bez utraty ogólności metody używać do określenia relatywnej siły relacji wartości z przedziału $\langle -1; 1 \rangle$.

$$W = \begin{Bmatrix} w_{11} & w_{1j} \\ \dots & \dots \\ w_{j1} & w_{jj} \end{Bmatrix}.$$

Macierz ta ma specyficzną budowę: elementy na przekątnej ($w_{x,x}, 1 \leq x \leq j$) mają wartość zero (uznajemy, że pojęcie nie może być poprzednikiem siebie samego), górny trójkąt macierzy W (z wyjątkiem przekątnej) zawiera siłę relacji dla par pojęć, natomiast dolny trójkąt jest odwróceniem tej relacji.

Wynikiem transformacji relacji między pojęciami ma być macierz kwadratowa D – macierz obrazująca relację zależności między dokumentami, o budowie podobnej do macierzy W . Wiersze i kolumny tej macierzy odpowiadają dokumentom, natomiast elementy d_{ij} zawierają względną wyliczoną siłę relacji poprzednictwa między dokumentem i a dokumentem j .

$$D = \begin{Bmatrix} d_{11} & d_{1k} \\ \dots & \dots \\ d_{k1} & d_{kk} \end{Bmatrix}.$$

Pokażemy, jak obliczyć pojedynczy element macierzy D . Przyjmijmy, że chcemy zbadać siłę relacji między dwoma dokumentami (*dokument 1* i *dokument 2*). W dokumentach tych występują tylko dwa pojęcia (*pojęcie 1* i *pojęcie 2*). Częstotliwość występowania tych pojęć w dokumentach została obliczona i zawiera się w następującej macierzy F

$$F = \begin{Bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{Bmatrix}.$$

Metodyk określił siłę zależności między *pojęciem 1* a *pojęciem 2* na 1,0. Oznacza to że uznał *pojęcie 1* za pierwotne względem *pojęcia 2*. Wobec tego macierz W (siły relacji między pojęciami) ma postać:

$$W = \begin{Bmatrix} 0 & 1 \\ -1 & 0 \end{Bmatrix}.$$

Jak obliczyć siłę relacji między dokumentami (czyli wyznaczyć macierz D)?

W tym celu należy wziąć pod uwagę:

- pozytywny wpływ na siłę relacji między dokumentami, na jaki może wskazywać częstotliwość wystąpienia *pojęcia 1* w *dokumentcie 1* i *pojęcia 2* w *dokumentcie 2*. Oznacza to, że *dokument 1* traktuje głównie o *pojęciu 1* (sugeruje to jego „łatwość” wobec „łatwości” *pojęcia 1*) podczas gdy *dokument 2* traktuje głównie o *pojęciu 2* (sugeruje to jego „trudność”). Wartość iloczynu częstotliwości i siły relacji między pojęciami $f_{11}f_{22}w_{12}$ bliska 1,0 wskazuje na to, że pojęcie ocenione jako „prostsze” występuje częściej w *dokumentcie 1* przy równocześnie częstym występowaniu *pojęcia „trudniejszego”* w *dokumentcie 2*;
- negatywny wpływ na siłę relacji między dokumentami, na jaki może wskazywać częstotliwość wystąpienia *pojęcia 1* w *dokumentcie 2* oraz *pojęcia 2* w *dokumentcie 1*. Wysoka wartość iloczynu (bliska 1,0) mogłaby sugerować, że *dokument 2* jest dokumentem poprzednikiem, ponieważ częściej występuje w nim pojęcie „łatwiejsze”, podczas gdy w *dokumentcie 1* często występuje pojęcie „trudniejsze”.

Uogólniając, wyznaczenie siły związku między dokumentami można uznać za zastosowanie formy dwuliniowej postaci:

$$d_{mn} = \sum_{i,j=1}^k w_{ij} f_{im} f_{jn},$$

w której f_{im}, f_{jn} to odpowiednio częstotliwości wystąpienia pojęć i, j w dokumentach m i n ; w_{ij} jest siłą relacji między pojęciami natomiast d_{mn} to siła relacji między dokumentami m i n . Dla wszystkich możliwych powiązań między dokumentami powyższy zapis można przekształcić w formułę:

$$D = F^T W F,$$

gdzie: D – macierz siły relacji między dokumentami,

F^T – transponowana macierz częstotliwości wystąpień pojęć w dokumentach,

W – macierz siły relacji między pojęciami.

2.5. Wyznaczanie relacji między pojęciami

W sposób analogiczny jak opisany powyżej, można oszacować siłę powiązań między pojęciami, jeżeli dana jest macierz siły relacji między dokumentami stanowiącymi szkolenia lub np. moduły szkoleń. Przesłanki wpływające na siłę relacji między pojęciami są analogiczne do tych opisanych powyżej, tak więc formuła obliczania macierzy relacji między słowami wygląda bardzo podobnie:

$$W = F^T D F$$

Stosując opisaną metodę, przeprowadziliśmy wstępne eksperymenty i przeanalizowaliśmy wybrane szkolenie z uwzględnieniem modułów w nim występujących jako pojedynczych dokumentów. Pojęcia wybraliśmy na podstawie nie opisywanej w artykule metody, biorąc pod uwagę rozkład częstości ich występowania w dokumentach. Wyniki eksperymentu – obliczone siły relacji między pojęciami oraz między dokumentami (modułami szkolenia) prezentuje rys. 2. Już pobieżny przegląd tak wczesnego eksperymentu sugeruje, że opisana powyżej metoda może być przydatna we wspomaganiu tworzenia sekwencji szkoleń.

3. Uwagi końcowe

Opisana w niniejszym artykule metoda wspomagająca sekwencjonowanie szkoleń oraz pojęć, mimo obiecujących wyników wstępnych, wymaga dalszych badań, a zwłaszcza eksperymentów na większym korpusie obiektów uczących. Także sam sposób wyznaczania sił relacji wymaga dalszych studiów. Na szczególną uwagę zasługuje zaobserwowany brak symetrii w zastosowaniu obu podejść metody, tzn. wyznaczenie siły relacji między dokumentami, a następnie wykorzystanie wyniku i wyznaczenie na powrót wyjściowej relacji między pojęciami nie prowadzi do początkowych danych. Może to sugerować, że na relację między dokumentami mają wpływ nie tylko wykorzystane w nich pojęcia, ale również takie czynniki, jak zastosowane przez twórcę sformułowania czy metoda dydaktyczna. Stanowi to przesłankę do udoskonalania metody w toku dalszych badań.

Literatura

- [1] Hübscher R. (2000), *Logically Optimal Curriculum Sequences for Adaptive Hypermedia Systems*, Adaptive Hypermedia and Adaptive Web-Based Systems: International Conference, AH 2000, Trento, Italy, August 2000. Proceedings. P. Brusilovsky, O. Stock and C. Strapparava, Springer-Verlag. 1147: 121-132.
- [2] Hübscher R. (2001), *What's in a Prerequisite*, Proceedings of International Conference on Advanced Learning Technology, Washington, IEEE Computer Society.
- [3] Lin J. i in. (2001), *On Workflow Enabled e-Learning Services*, Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges. T. Okamoto, R. Hartley, Kinshuk and J.P. Klus. Madison, USA, IEEE Computer Society: 349-352.
- [4] Shang Y. i in. (2001), *An Intelligent Distributed Environment for Active Learning*, "Journal of Educational Resources in Computing" 1(2): 17.
- [5] Specht M., Oppermann R. (1998), *ACE - Adaptive Courseware Environment*, "The New Review of Hypermedia and Multimedia" 4(1): 141-161.
- [6] Wiley D.A. (2000), *Learning Object Design and Sequencing Theory*, Department of Instructional Psychology and Technology, Brigham Young University: 11-131.

LEARNING OBJECT SEQUENCING

Summary

Different e-learning technologies use learning objects as knowledge resources. The methodologies of designing learning objects focus on creating single learning objects. Most of them neglect the representation of dependencies between learning objects. The research pursued on this topic led to several solutions of the problem. None of them however is well suited for distance education in the Internet environment. Such a solution is necessary to develop methods of automatic sequencing of learning objects. In this article the authors propose a solution based on directed acyclic graphs and graph algorithms for creating such learning paths, that can be easily adopted for distance learning on the Internet. Learning objects and the dependencies between them are represented as paths in the graph. This enables reasoning based on the graph structure about the learning paths adapted for specific user requirements. The authors propose also a method of semi-automatic creation of dependencies between learning objects grounding on the knowledge extracted from the learning objects.