Keywords linking method for selecting educational web resources à la ZigZag

Mirjana Andric*

Intelligence, Agents, Multimedia Group,
School of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ, UK
Fax: +44-23-8059-2865
E-mail: mira@computer.org
*Corresponding author

Vladan Devedzic

FON – School of Business Administration,
University of Belgrade,
Jove Ilica 154, P.O. Box 52,
Belgrade 11000, Serbia
E-mail: devedzic@etf.bg.ac.yu

Wendy Hall and Leslie Carr

Intelligence, Agents, Multimedia Group,
School of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ, UK
E-mail: wh@ecs.soton.ac.uk
E-mail: lac@ecs.soton.ac.uk

Abstract: The authors of the web-based courseware typically face problems such as how to locate, select and semantically relate suitable learning resources. This paper proposes a tool that supports the authors in their tasks of selection and grouping the learning material. The 'à la' (Associative Linking of Attributes) in Education, enhances the search engine results by extracting the attributes (keywords and document formats) from the text. The relationships between the attributes are established and visualised in a novel hypertext paradigm using the ZigZag principles. Browsing the related metadata provides a quick summary of the document and can hence help in faster determining its relevancy. Also, the proposed solution enables better understanding of why some resources are grouped together as well as providing suggestions for the further searches. The results of a user trial indicate high levels of user satisfaction and effectiveness.

Keywords: authoring support tools; extracting metadata; hypertext structures; web-based teaching; ZigZag; zzstructures.

Reference to this paper should be made as follows: Andric, M., Devedzic, V., Hall, W. and Carr, L. (2007) 'Keywords linking method for selecting educational web resources à la ZigZag', *Int. J. of Knowledge and Learning*, Vol. 3, No. 1, pp.30–45.

Biographical notes: Mirjana Andric is a PhD student at the University of Southampton, UK. She has received a BS and an MS degrees from the School of Electrical Engineering University of Belgrade, Serbia and Montenegro. Her current research includes hypertext structures, metadata, recommendation systems and linking and navigation.

Vladan Devedzic is currently a Professor of Computer Science at the University of Belgrade, Serbia. He received a BS, an MS and a PhD from the School of Electrical Engineering, University of Belgrade, Serbia. His current efforts and research interests are oriented towards practical engineering aspects of developing intelligent educational systems on the Web. He is a member of the *International Society for Artificial Intelligence in Education* (*AIED*).

Wendy Hall is a Professor of Computer Science at the University of Southampton, UK. She was the Founding Head of the Intelligence, Agents, Multimedia (IAM) Research Group in the School of Electronics and Computer Science at Southampton and is currently the Head of the School. Her research interests include the development of web technologies (particularly the semantic web), open hypermedia systems and link services, advanced knowledge technologies, digital libraries, multimedia information management, agent-based systems and human–computer interaction. She was the President of the British Computer Society (2003–2004) and was awarded a CBE in 2000.

Leslie Carr is a Senior Lecturer in the IAM group at Southampton University where he coordinates the various EPrints research and development activities. His research interests include the application of hypermedia technologies to digital libraries and the deployment of engineering methodologies to produce effective and reliable information systems. His work is centred around improving human communication through electronic means, including XML, RDF, ontologies and distributed information environments. He is a member of the OAI technical committee and director of Eprints software.

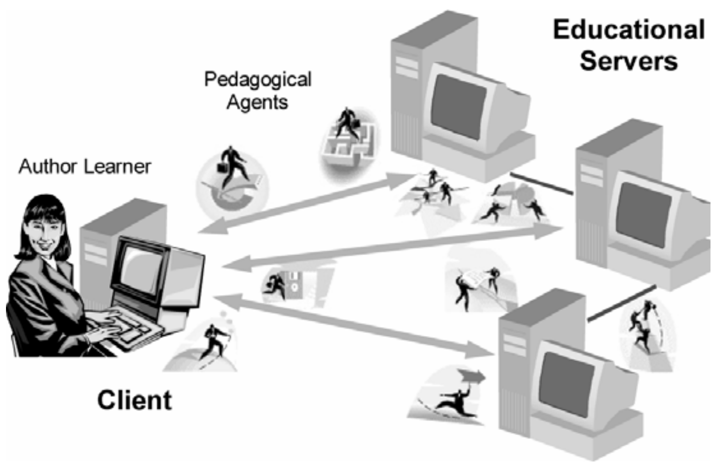
1 Introduction

1.1 Authoring web-based courseware

Web-based education has become a very important branch of educational technology. For learners, it provides access to information and knowledge sources that are practically unlimited, enabling a number of opportunities for personalised learning, tele-learning, distance-learning and collaboration, with clear advantages of classroom independence and platform independence. On the other hand, teachers and authors of educational material can use numerous possibilities for web-based course offerings and teleteaching, including authoring tools for developing web-based courseware. Furthermore, cheap and efficient storage and distribution of course materials, hyperlinks to suggested readings, digital libraries and other sources of references relevant for the course are readily available.

In the context of web-based education, educational material is generally distributed over a number of *educational servers*, Figure 1 (Devedzic, 2003). The authors (teachers) create, store, modify and update the material working with an *authoring tool* on the client side.

Figure 1 The context of the web-based education



Source: After Devedzic (2003).

Pedagogical agents provide the necessary infrastructure for knowledge and information flow between the clients and the servers. On behalf of the learners, they access educational content on the servers by using high-level educational services. Educational content is any educational material pedagogically organised and structured in such a way that interested learners can use to get introduced to a knowledge domain, deepen their understanding of that domain and practice the related problem-solving skills.

In a typical scenario of creating learning material in such a context, the author would browse a number of educational servers and look for the other resources on the web. Then (s)he would reuse and reorganise parts of the material found, creating a new learning material. Generally, the new material will take the form of a sequence or a network of interconnected learning objects. Some typical problems that may arise in this scenario are:

- How to locate suitable learning resources on the web?
- How to select the most appropriate resources for further reuse in composing the new learning material to suite the learners' needs?
- How to effectively correlate selected resources and create groups of semantically related resources to be used in the next step of creating the new material?

With the current technology, the author typically uses a search engine to locate the learning material on the web. One drawback is that it is a keywords-based search, since the metadata by which the educational content on the web can be classified is still largely lacking. Although there have been some advances to this end in the area of the semantic web (Devedzic, 2004), it is not commonplace yet. Moreover, in order to select a resource (find out whether it is relevant or not), the author must read it through. If (s)he prefers to store the reference to the resource for future use, it results in individual bookmarking and creates another typical classification problem – to remember what web pages were similar and for what reason.

1.2 Summary of the proposed solution

The solution to the stated problem, proposed in this paper, builds on top of the existing solution consisting of search engine usage and hypertextually connected metadata. In our 'à la' (Associative Linking of Attributes) method (Andric et al., 2004, 2005), the results obtained using the search engines solution are enhanced by post-processing. In essence, search engine results are retrieved and the attributes, mainly keywords, are extracted from the textual resources. Then, the relationships between the attributes are statistically analysed and established. Attribute associative links, that is, instantiations of a semantic relationship between information elements (Lowe and Hall, 1999), are then woven into a hypertext structure. Subsequently, the attribute connections are visualised in a novel hypertext paradigm using the ZigZag (Nelson, 1998) principles. The author is then able to browse the keywords and their links and to select the most promising documents. Finally, selected documents and their keywords are saved into a document collection, ready for later browsing and amending. This solution seems to be more effective than purely the use of a search engine because:

- it enables better understanding of why the resources are similar, that is, which keywords they share
- it provides a set of keywords acting as a summary of the web document, which enables easier selection of the relevant keywords
- finally, it provides suggestions of keywords to further search by.

The prototype system was built in order to investigate the research ideas. The system was evaluated in a user trial in which a set of 20 teachers were trying to sequence a web-based course with and without the 'à la' system. The results obtained using a post-trial questionnaire and the Wilcoxon statistical test, indicate the higher level of user satisfaction and effectiveness, compared to the standard, search-engine only, solution.

2 Background

2.1 Ontologies and the semantic web in education

An important related research area in education is that of ontologies and ontology-aware authoring tools (Aroyo and Dicheva, 2004; Aroyo and Mizoguchi, 2003; Devedzic, 2004). Ontologies enable machine understanding and machine processing of different contents. In education, they standardise and provide interpretations for educational contents on the web. However, making the contents understandable and processable by pedagogical agents requires the corresponding educational web pages to contain semantic markup, that is, descriptions which use the terminology that one or more ontologies define and contain pointers to the network of ontologies. Using ontologies as references in marking-up educational pages on the semantic web enables knowledge-based indexing and retrieval of services by pedagogical agents and humans alike, as well as automated reasoning about the services, such as how to use them, what parameters to supply, what results to expect and so on. Artificial Intelligence (AI) techniques are used for ontology creation. The most notable classical work in the AI in Education community related to the development of educational ontologies comes from the Mizoguchi Lab at Osaka University, Japan (e.g. see Mizoguchi and Bourdeau, 2000 and from Murray, 1998). Work of Gasevic and Hatala (forthcoming) shows how different ontologies can be mapped to each other in order to support querying of remote educational servers using the terms from the local educational ontology.

Ontology-aware authoring tools assist the authors in structuring the domain, the courseware and the topics to be presented to the learner. They use domain ontologies as firm and stable knowledge backbone to which the author can attach the courseware material. Using ontology-based layers in the courseware authoring architecture is described by Aroyo et al. (2002). Domain visualisation techniques help ontology-aware authoring tools communicate ontological knowledge to the author more effectively. Moreover, these tools guide the authoring process by providing the ontology of authoring tasks.

The only problem with ontologies is that they are still relatively sparse. A vast majority of domains and topics are still not supported by appropriate ontologies, which makes it difficult for authors to use ontological support.

2.2 Web mining

Another growing branch of related research is web mining for learning resources (e.g. see Trausan-Matu et al., 2002). Web mining is the process of discovering potentially useful and previously unknown information and knowledge from web data (Cooley et al., 1997). It encompasses tasks such as automatic resource discovery, automatic extraction and preprocessing of desired data from web documents, discovery of common patterns across different websites and validation and/or interpretation of discovered patterns (Chakrabarti et al., 1999).

The area of web mining relevant for the topic of this paper is called web content mining. It refers to deploying personalised, ontology-enabled pedagogical agents to continuously go collect globally distributed content and knowledge from the web (large web data repositories such as documents, logs and services) and organise it into educational web servers (Devedzic, 2004). The collected data can then be incorporated with locally operational knowledge bases and databases to provide a dedicated community of learners with centralised, adaptable, intelligent web services. A pioneering work in this direction is presented by Trausan-Matu et al. (2002). The idea is that the knowledge the learners need to learn is not static but changes dynamically due to the continuous development and change of available resources on the web, hence any sequencing of the learning material in a web-based intelligent educational system should reflect that dynamics accordingly.

2.3 ZigZag

ZigZag represents an innovative information storing paradigm introduced by the hypertext pioneer (Nelson, 1998, 2003). Its main concept of using a complex matrix-like structure for storing and manipulating pieces of information in multiple contexts "may be thought of as a multidimensional generalization of rows and columns, without any shape or structure imposed" (Nelson, 1998). The data stored in the elementary cells of this matrix-like structure, known as *zzstructure*, are connected using untyped links. The content of a cell can be a simple piece or text or it can be any more complex object. The cells can be connected to each other along an unlimited number of dimensions, which effectively represent types of relationships. It is convenient to colour the links belonging to the same dimension with the same colour.

However, the way of connecting cells in the zzstructure conforms to a limitation, known as a *restriction R* (McGuffin and Schraefel, 2004). A cell can participate in a dimension by connecting to the other cell(s) (or in the special case to itself) via two poles, positive and negative. A restriction *R* says that for a particular dimension, a cell can be connected in such a way that none, one or both of the poles are used. Therefore, there exists a constraint that a cell can have at most one neighbour on each side. In other words, cells are connected in a series of sequences or lists if only one dimension is observed at a time. No one-to-many links are allowed (Moore and Brailsford, 2004). This collection of strands or paths, in zzstructure called ranks, form a network, effectively, an edge-coloured directed multigraph subject to *restriction R*, explained above (McGuffin and Schraefel, 2004).

Although zzstructure links possess a direction, starting at the positive side of one cell and finishing in the negative side of another, they can be traversed both ways. Loops are allowed. The furthest cell in the negward direction of each rank (i.e. not a closed loop) is called the headcell. The headcell, in a case of the vertical rank, would be the cell at the top.

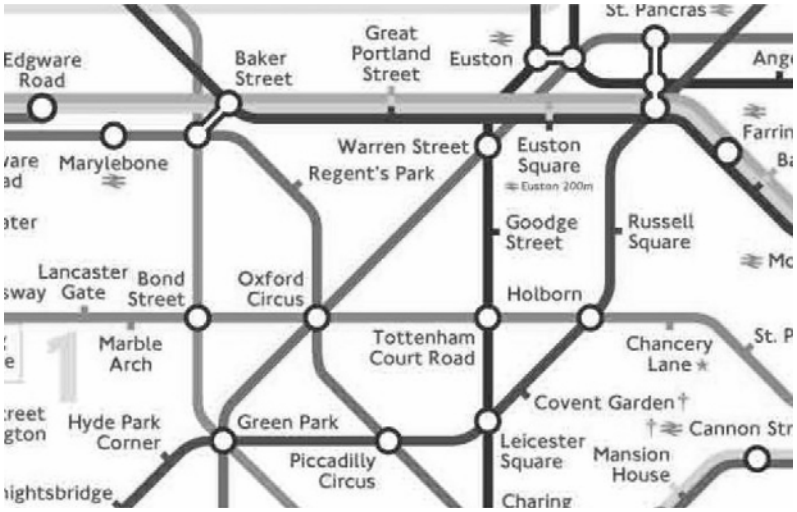
Cells do not repeat within the structure, but if it is necessary for a cell to participate in a one-to-many relationship, a clone of the cell can be created. In this case, a virtual repetition of the cell content is achieved, in order to enable the cell to be connected to many other cells via the same dimension, without breaking the ZigZag restrictions.

The zzstructure principle allows for an interesting effect of criss-crossing lists, where a cell can exist on many lists at the same time (Nelson, 1998). The structure is in general extremely difficult to visualise. Usually, only a portion of the whole structure is shown at a time, typically revealing cells and links in 2 or 3 dimensions.

2.4 London tube: zzstructure example

An excellent example of a zzstructure is the system of London underground (http://www.thetube.com) train lines and stations. Stations represent cells while the train lines can be considered as dimensions. Some stations can belong to more than one line, where different ranks intersect. Moreover, in the example of the London tube system given in Figure 2, each line is given a name and a specific colour. A traveller on the network can follow some route (rank) or change the line (dimension) on a certain station (cell), providing that such cell offers a choice of interconnection.

Figure 2 Portion of the London underground network on a map (London underground)



The diagrams in Figures 3 and 4 provide a view on the zzstructure using a Southampton University developed ZigZag Browser (Carr, 2001). The cell in a ZigZag browser is represented by a rectangle, while links are represented as arrows. As can be seen in Figures 3 and 4, some cells have several links indicated with slanted arrows. A traveller at the Piccadilly Circus station can decide to continue right, following the blue-coloured, Piccadilly line towards Leicester Square or to change the dimension/line to the purple-coloured, Bakerloo Line and go up to Oxford Circus as shown in Figure 4. As the user moves in the multidimensional space of the train lines, the currently accessible cells and links are revealed, while the inaccessible ones become invisible.

Figure 3 Portion of the London underground network in the ZigZag browser (Carr, 2001) – current station Piccadilly Circus

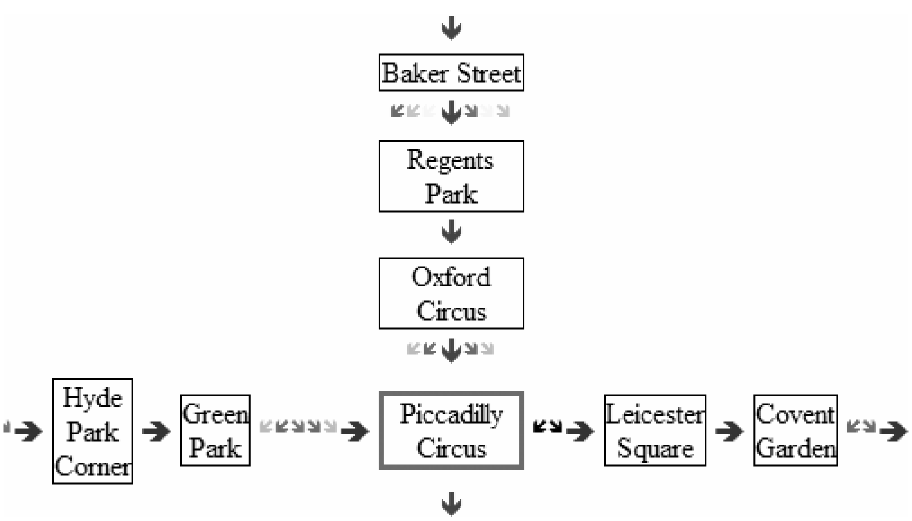
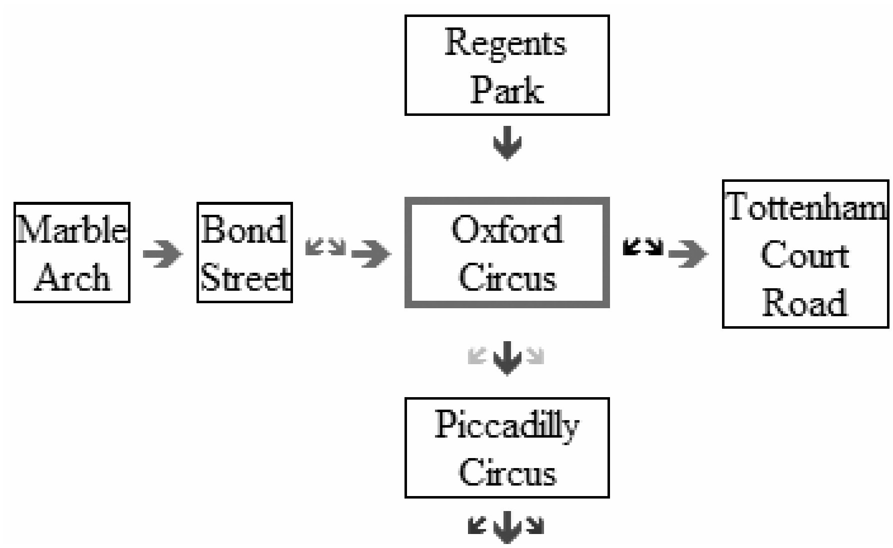


Figure 4 Portion of the London underground network in the ZigZag Browser (Carr, 2001) – navigating vertically to the station Oxford Circus



Two examples of ZigZag applications to the real-world domains are described by Moore et al. (2004) and Moore and Brailsford (2004): zzPhone, the information manager for mobile phones and the Bioinformatics workbench for creating and manipulating interconnected biological information such as atoms in the metabolic Krebs Cycle.

3 The 'à la' platform for education

3.1 System overview

The idea of presenting the interconnected pieces of the information, in fact the simple ontology network, in zzstructures, has been an inspiration for the 'à la' system (Andric et al., 2004, 2005). The central idea of the 'à la' method for education is that extracting some metadata (or attributes) from the web textual resources, analysing their relationships and storing them into a browsable zzstructure, can improve the process of searching and selecting the learning material on the web.

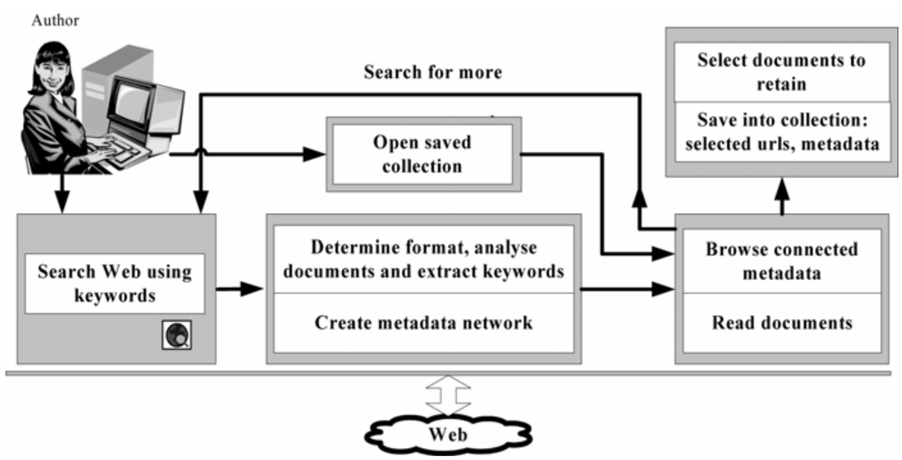
In order to achieve the set goal, the system needs to perform the three main steps:

- build the attributes-links network
- provide the user with a browser tool for this network
- select and save references (URLs) and attributes of chosen web documents.

The 'à la' platform for education architecture is presented in Figure 5. The course author can use this enriched search system either by posting a regular query to a search engine or opening a previously saved preprocessed document collection. In the first case, a set of keywords is sent to a search engine (e.g. Google) and the results analysed. Two types of attributes are harvested: a file format and a set of keywords, using a TF-IDF machine learning technique from the 'Linking in Context' project (El-Beltagy et al., 2001). Then, the algorithm for creating the metadata network builds an attribute network and stores it to a zzstructure, which is later presented to the user. In the second case the user opens the attribute network previously saved in a collection. The user can then browse the attribute network, familiarising her/himself with the keywords, formats and with the information about which ones of them appear in which documents. From that moment, the user can:

- decide to read the content of some document if its keywords or links to other documents appear to be of the interest
- decide to use the browsed keywords in order to expand or replace the old search terms and then ask for more search engine results
- select the interesting documents and save the whole structure in a named document collection for the later usage.

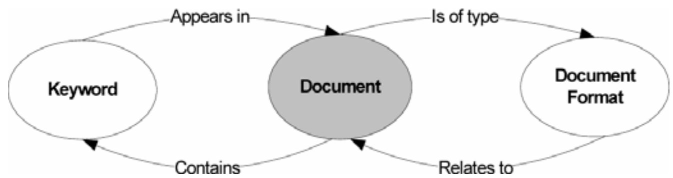
Figure 5 The 'à la' platform for education: lock architecture



3.2 *The 'à la' implementation highlights*

The 'à la' method for education uses a very simple set of attributes and relationships for building its metadata network. Only two types of metadata are considered: the web *document format* (such as HTM or PDF) and the *keyword*, meaning the term that is among the most frequent terms in the text. This set of metadata is chosen because it is available on the web in most of the cases. The attributes and the relationships in which they participate are shown in Figure 6. Note that there exists a relationship for each direction, as for example a document can contain many keywords, while a keyword can appear in many documents.

Figure 6 Types of attribute links, that is, relationships analysed in the 'à la' system



Therefore four relationships can be analysed and mapped to the zzstructure dimensions:

- document – keywords
- keyword – documents
- document – document formats
- document format – documents.

In our zzstructure building algorithm relationship types get mapped to dimensions, attribute values become cells and for example keywords of the same documents get connected by links.

Following the common TF-IDF approach, the search engine results are analysed for keywords. Document formats are determined first. Text parts are considered and keywords extracted, stemmed and ordered by frequency. Only the most frequent keywords are taken, the rest are pruned. The frequencies are normalised to calculate weights of each attribute in the resulting document feature vector. N-tuples *Document* (document identifier) – *Attribute* (attribute type) – *Value* (attribute instance) – *Weight* are formed. The network creation starts with the cells creation. The attribute values, the actual instances of the keywords, document titles or document formats, become unique cells in a zzstructure. Each of the four relationships becomes a dimension. Then, the dimensions are iterated and for each of the dimension all its ranks built.

The algorithm for populating the metadata network takes document attributes from the 'Document-Attribute-Value-Weight' pool, analyses them using the information to which document they belong and establishes attribute links. The algorithm creates a rank for a given dimension using a first attribute instance, for example, Document (title of a document) as a headcell. Instances of a second attribute type, for example, Keyword, which are sharing the same document as the first attribute value, are then ordered by decreasing weights and woven into a rank. Since ranks on the same dimension are not allowed to intersect, clone cells are used when necessary.

The actual rank in the example 'diet' related websites network could look like this: Atkins Home–atkins–nutrition–carb (dimension *Document Contains Keywords*). On the other side the (stemmed) keyword 'nutrition' could have its own rank in the dimension *Keyword Appears in Documents*: nutrition–DietSite–Atkins Home.

The example of the zzstructure in XML is given in Figure 7. It shows a portion of one zzstructure created by the algorithm. For example, the entry *<dimension name="Document Contains Keywords" description="Doc-Kw" color="Red"/>* is created by providing the identifier – the dimension name, colour from a list of HTML colours and the description as a shortened name *Doc-Kw*. The <cells> entries show the interconnections of cells. For example, cell number 3, 'nutrition' has posward connections towards the other keyword in the dimension 'Doc-Kw' and another posward connection towards the cell representing a document in the dimension 'Kw-Doc'.

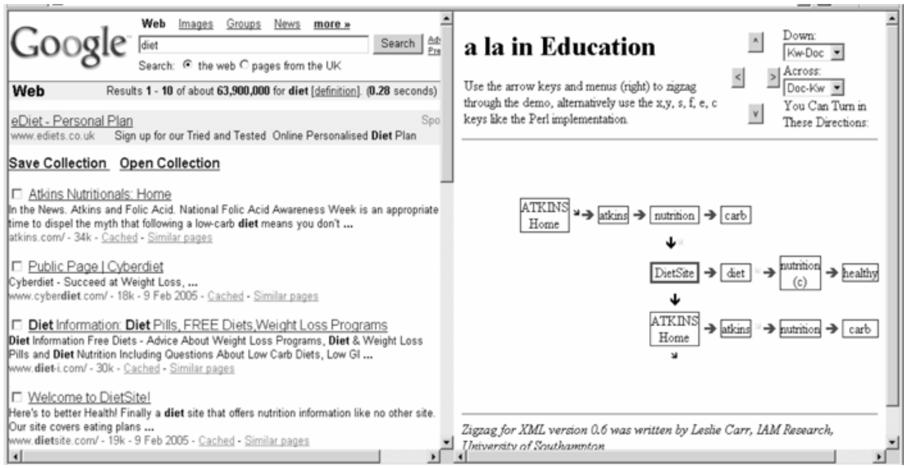
Figure 7 Example of a zzstructure in XML

```
<!DOCTYPE zigzag SYSTEM "zigzag.dtd">
<?xml version="1.0"?>
<?xml:stylesheet type="text/xsl" href="zigzag.xsl"?>
<zigzag>
<title>a la in Education</title>

<dimensions>
<dimension name="Document contains Keywords" description="Doc-Kw" color="Red"/>
<dimension name="Keyword appears in Documents" description="Kw-Doc" color="Black"/>
<dimension name="Document has Formats" description="Doc-Fmt" color="Blue"/>
<dimension name="Format relates to Documents" description="Fmt-Doc" color="Magenta"/>
<dimension name="Clone" description="Clone" color="Yellow"/>
</dimensions>
<cells>

<cell n="1">
      <content>ATKINS Home</content>
      <link direction="Document contains Keywords" posward="2"/>
      <link direction="Document has Formats" posward="17"/>
</cell>
<cell n="2">
      <content>atkins</content>
      <link direction="Document contains Keywords" posward="3"/>
</cell>
<cell n="3">
      <content>nutrition</content>
      <link direction="Document contains Keywords" posward="4"/>
      <link direction="Keyword appears in Documents" posward="9"/>
</cell>
…
<cell n="9">
      <content>DietSite</content>
      <link direction="Keyword appears in Documents" posward="1"/> …
</cell>
…
</cells>
</zigzag>
```

4 User interaction

In the example of the user interaction with the system (Figure 8), the course author wants to select material for the guided tour around ‘diet’ devoted websites. The author enters the term ‘diet’ into the search box and initiates processing of the results. The page of the prototype system is divided into two areas. The left side resembles the search engine result with the addition of the selection capability of the interesting documents for saving in the collection.

Figure 8 User interaction example in the 'à la' system for education



The user can browse a network of cells in a zzstructure in the right side pane and (s)he has two dimensions to choose at a time: Across and Down. Navigation starts at the current cell, keyword 'diet', which is specially marked. If a dimension which has links towards the current cell is selected, the connected cells will be shown as arrays in horizontal or vertical ranks. The user can then navigate the ranks up/down or left/right. Whenever the user changes the current cell, the zzstructure view might change: some new cells might be revealed, some old hidden, all depending on the current position and the two selected dimensions. When a dimension is changed, the new one will replace the old one and the view will change accordingly.

The user can see that this particular keyword appears on the Websites 'DietSite'. The user then decides to navigate to a 'DietSite' and select vertically 'Kw-Doc' and horizontally 'Doc-Kw' dimensions. The vertical dimension allows him/her to note that keyword 'nutrition' appears on two sites 'DietSite' and 'Atkins Home'. The horizontal dimension shows for the two visible sites their keywords. It is interesting to note a multiple appearance of the term 'nutrition'. From it we can deduce the following:

- In the first horizontal rank (*Atkins Home-atkins-nutrition-carb*) we see that the term 'nutrition' appears among keywords in the 'Atkins home' website as the second frequent for that website.
- By following the vertical rank (*nutrition-DietSite-Atkins Home*) we find out that this term has two websites where it is a keyword.
- By looking at the second horizontal rank we find out that 'nutrition (c)' appears as the clone of the original 'nutrition' term. The original and a clone cells are connected via a Clone dimension indicated by a yellow arrow. The reason for cloning in this case is the fact that the same cell would need to be involved in two different ranks of the same dimension, which goes against the ZigZag rules.

5 System evaluation

A set of 20 teachers was selected for the evaluation. The assumption made was that the teachers were reasonably and equally skilled in the internet search techniques and that they use them regularly. The users were randomly divided into two equal groups. The first group was given a task to select material for the course in their own area, using strictly a search engine and the bookmarking techniques. After a brief demonstration, the second group was instructed to perform the same task but using the 'à la' tool. The groups were then switched. The duration of the sessions was limited to 1 hour. After that, they were presented with the following questionnaire for each of the systems.

Provide a grade from 1 (the worst) to 10 (the best) for each of the following questions (Table 1):

- How easy was it to learn to use the system?
- How friendly was the user interface?
- How effective was the system in supporting your task?
- What was the overall satisfaction with the system?

The Wilcoxon signed rank test was used to compare the obtained results, in order to show the differences between the paired observations.

Table 1 Evaluation results showing comparison to the classical approach using ranking 1 to 10

Metrics used	*Avg. rank (search engine)*	*Avg. rank ('à la' method)*	*No. of <> pairs*	*Probability of identical distribution*
Method learnability	7.70	6.75	17	≤0.06487
Friendliness of the user interface	8.00	6.70	14	≤0.01074
Effectiveness	7.30	8.40	15	≤0.03534
Overall user satisfaction	7.90	8.25	13	≤0.41430

The results indicate that the initial learnability and friendliness of the user interface are lower for the 'à la' system compared to the classical solution. However, this observation is expected as the way of using the standard search engine solution is widely known. On average, the results demonstrate better effectiveness and the overall satisfaction with the 'à la' system for education. On the other hand, the future work should explore the larger user population and the usage of other metrics, in order to confirm and expand the observations obtained in this trial. This is especially related to the effectiveness which should be objectively measured.

6 Related work

There are numerous examples of research undertaken in extracting the keywords from the textual resources, especially in the content-based recommender systems, such as Balabanovic and Shoham (1997) and Mladenic (1999).

The ‘à la’ system in Education combines that approach with overlaying the extracted metadata network over a document collection space, using a concept similar to Bruza’s ‘hyperindices’. A semantic index space, ‘Hyperindices’, as a set of browsable concept descriptors is proposed in Bruza’s two-level architecture for hypertext documents in Bruza (1990). The user can alternatively navigate in the index and the underlying content space or hyperbase, by ‘beaming up and down’.

The project ‘Linking in Context’ (El-Beltagy et al., 2001) analyses the content of searched and downloaded web pages in order to find similar ones and ultimately reuse manually authored hyperlinks. The ‘à la’ prototype incorporates the keyword extracting module developed in this project.

Note that zzstructures are conceptually similar to semantic networks, a traditional knowledge representation structure in AI (Quillian, 1968). In semantic networks, the nodes – concepts – are connected by arrows – binary relations between the concepts where appropriate. Pretty much like a cell in a zzstructure, a node in a semantic net can be connected to any number of other nodes and the links can denote different relationships between nodes. However, there is no notion of a dimension or a rank in semantic networks, at least not an explicit and intentional/a priori one. Unlike zzstructures, they were not developed with criss-crossing in mind and do not origin from hypertext research.

Similarly to “à la” in Education approach, graphs with semantic meaning are used for organisation of the educational material (Mittal et al., 2003).

On the topic of visualising complex structures of interconnected keywords, the existing research is mostly focused on representing hierarchical structures. The Anacubis demo (Anacubis Books Demo, http://www.anacubis.com/amazondemo/amazon/) of Amazon/Google visual search for books, shows only a portion of the structure, centralised on the found item (book in this case). It represents different relationships, such as *Related*, *Author* and *Also Bought*, in different colours. However, the items and relationships cannot be easily navigated and only the immediate neighbourhood, that is, items directly linked, is displayed. As a comparison, visualisation in ‘à la’ zzstructure provides better navigation features and more flexible relationship display as it is not limited to showing only hierarchical structures.

7 Conclusion

Teachers and authors developing web-based courseware typically face problems in locating and organising suitable learning resources. They resort to keyword-based search using searching engines and the bookmarking techniques. The ‘à la’ (Associative Linking of Attributes) in education, presented in this paper, offers methods for improving the classical approach to the problem of authoring web-based courses. The ‘à la’ technique consists of enhancing the search engine-based solution in the following way:

- textual documents from the search results and two types of extracted attributes (keywords and file formats) are analysed
- relationships between attribute instances are statistically analysed and the most frequent ones established
- attribute links are presented to the user in a browsable hypertext structure using ZigZag principles.

In order to evaluate the mentioned research ideas, the 'à la' in education prototype was implemented and evaluated during a user trial. The user study looked into how easy it was to learn to use the system, how friendly the interface was, how effective was the system in supporting the user's task and finally, what the overall user satisfaction was. The system was compared with the classical solutions of using only the search engine. A group of teachers was asked to locate and select suitable web resources for a web course. The aim of the trial was to confirm the expected solution contributions:

- browsing the related metadata (keywords and formats) along the search results helps determine the relevancy faster by offering a sort of quick summary of the document
- shared keywords help establish which documents could be semantically related
- extracted keywords can provide suggestions for further searching.

Results indicated that, after the initial learning effort, the 'à la' prototype showed potential to have a high level of effectiveness and a better overall user satisfaction.

The use of a system by a group of teachers opens up a new research direction: the possibility of utilising the system in a collaborative environment. Ideas about sharing the authoring experiences also raise personalisation issues; therefore possible future work might investigate the use of a personalised, continuous web content mining agent.

Acknowledgement

We would like to thank our study participants from the University of Southampton, UK and the University of Belgrade, Serbia and Montenegro. 'ZigZag' is a trademark of Dr. T. Nelson, Oxford Internet Institute, UK.

References

Andric, M., Devedzic, V., Hall, W. and Carr, L. (2005) '"à la" in education: keywords linking method for selecting web resources', *Proceedings of International Conference on Artificial Intelligence in Education (AIED'05)*, July.

Andric, M., Hall, W. and Carr, L. (2004) 'Assisting artifact retrieval in software engineering projects', *Proceedings of the ACM Symposium on Document Engineering (DocEng)*, October, pp.48–50.

Aroyo, L. and Dicheva, D. (2004) 'The new challenges for e-learning: the educational semantic web', *Educational Technology and Society*, Vol. 7, No. 4, pp.59–69.

Aroyo, L., Dicheva, D. and Cristea, A. (2002) 'Ontological support for web courseware authoring', in S. Cerri, G. Gouarderes and F. Paraguacu (Eds). *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems ITS'2002*, June, pp.270–280.

Aroyo, L. and Mizoguchi, R. (2003) 'Authoring support framework for intelligent educational systems', *International Proceedings of Conference on Artificial Intelligence in Education (AIED'03)*, July, pp.362–364.

Balabanovic, M. and Shoham, Y. (1997) 'Fab: content-based, collaborative recommendation', *Communications of the ACM*, Vol. 40, No. 3, pp.66–72.

Bruza, P. (1990) 'Hyperindices: a novel aid for searching in hypermedia', *Proceedings of the ECHT '90*, pp.109–122.

Carr, L. (2001) 'ZigZag for web browsers', Available at: http://www.ecs.soton.ac.uk/~lac/zigzag.

Chakrabarti, S., Dom, B., Gibson, D., Kleinberg, J., Kumar, S., Raghavan, P., Rajagopalan, S. and Tomkins, A. (1999) 'Mining the link structure of the world wide web', *IEEE Computer*, Vol. 32, No. 8, pp.60–67.

Cooley, R., Mobasher, B. and Srivastava, J. (1997) 'Web mining: information and pattern discovery on the world wide web', *Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)*.

Devedzic, V. (2003) 'Key issues in next-generation web-based education', *IEEE Transactions on Systems, Man, and Cybernetics, Part C – Applications and Reviews*, Vol. 33, No. 3, pp.339–349.

Devedzic, V. (2004) 'Education and the semantic web', *International Journal of Artificial Intelligence in Education (IJAIED)*, Vol. 14, pp.39–65.

El-Beltagy, S., Hall, W., De Roure, D. and Carr, L. (2001) 'Linking in context', *Proceedings of the ACM Hypertext '01 Conference*, July, pp.151–160.

Gasevic, D. and Hatala, M. (forthcoming) 'Ontology mappings to improve learning resource search', *British Journal of Educational Technology*.

Lowe, D. and Hall, W. (1999) *Hypermedia and the Web, An Engineering Approach*, John Wiley & Sons.

McGuffin, M. and Schraefel, M. (2004) 'Hyperstructure: a comparison of hyperstructures: zzstructures, mSpaces, and polyarchies', *Proceedings of the ACM Hypertext '04 Conference*, August, pp.153–162.

Mittal, A., Dixit, S. and Maheshwari, L. (2003) 'Enhanced understanding and retrieval of e-learning documents through relational and conceptual graphs', *Workshop on Technologies for Electronic Documents for Supporting Learning Held in conjunction with the Artificial Intelligence in Education (AIED), 11th International Conference on Artificial Intelligence in Education*, July.

Mizoguchi, R. and Bourdeau, J. (2000) 'Using ontological engineering to overcome common AI-ED problems', *International Journal of Artificial Intelligence in Education*, Vol. 11, pp.1–12.

Mladenic, D. (1999) 'Text-learning and related intelligent agents: a survey', *IEEE Intelligent Systems*, pp.44–54.

Moore, A. and Brailsford T. (2004) 'Unified hyperstructures for bioinformatics: escaping the application prison', *Journal of Digital Information*, Vol. 5, No. 1, Article No. 254.

Moore, A., Goulding, J., Brailsford, T. and Ashman, H. (2004) 'Practical applitudes: case studies of applications of the ZigZag hypermedia system', *Fifteenth ACM Conference on Hypertext and Hypermedia*, Santa Cruz, CA.

Murray, T. (1998) 'Authoring knowledge-based tutors: tools for content, instructional strategy, student model, and interface design', *The Journal of the Learning Sciences*, Vol. 7, No. 1, pp.5–64.

Nelson, T. (1998) 'What's on my mind', *Invited Talk at The 1st Wearable Computer Conference*.

Nelson, T. (2003) 'Structure, tradition and possibility', *Proceedings of the ACM Hypertext '03 Conference*, August, p.1.

Quillian, M.R. (1968) 'Semantic memory', in M.L. Minsky (Ed). *Semantic Information Processing*, Cambridge, MA: MIT Press, pp.216–270.

Trausan-Matu, S., Maraschi, D. and Cerri, S. (2002) 'Ontology-centered personalized presentation of knowledge', *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, *ITS 2002*, June, pp.259–269.