Integrated Multimodal Copy-Paste Checking

Narayanan Kulathuramaiyer
Faculty of Computer Science and Information Technology,
University Malaysia Sarawak, Malaysia
Email: nara@fit.unimas.my

Bilal Zaka
Institute for Information Systems and Computer Media,
Graz University of Technology, Austria
Email: bzaka@iicm.edu

***Abstract*— Although the use of advanced computing and communications technology has made learning a significantly richer experience for learners, it also introduces greater ease in committing academic cheating. There is a growing increase of internet usage in educational communities. Powerful search engines and media gateways gives an ease of access to wide spreading and different types of information sources, this in turn reduces the need for learners to perform diligent research or study. A learner can easily copy information, found on the internet, relevant to the task at hand. E-learning systems would then need to incorporate extended functionalities to enable multimodal copy-paste checking. Besides the mere detection of text copy-paste activities, there is a need to determine deeper similarity relations in text and various other media formats. Furthermore, the introduction of a facility to monitor and check the student's ways of searching, reading and writing can improve the learning process. Our previous paper [1] describes the architecture and design of an integrated Copy-Paste system aimed to provide a platform addressing these concerns. This paper extends the work and shows our pioneering explorations in extending student modeling capability of e-learning systems. We discuss the use of layered similarity assessment techniques to add value to conventional copy-paste detection systems.**

***Index Terms*— Plagiarism, E-Learning, Intellectual Property Rights, Multimodal Copy-Paste, WritePrint, Similarity Assessment**

I. Introduction

During the past two decades, use of computer technology particularly World Wide Web by students, has become a customary habit. This behavior introduced a growing problem of cyber plagiarism in learners. Due to widespread use of web and digital libraries, access to educational and research contents is constantly improving. This increasingly open access to intellectual contents provides equally open and easy possibilities of copying or reusing of the contents. (whether it is intentional or un-intentional or done legitimately). According to a survey by center of academic integrity [2], 40% or more students admitted to plagiarism in year 2005 compared to 10% in year 1999. This figure is increasing every year. In another survey [3] done by the Association of Teachers and Lecturers (ATL) in the U.K., more the 50% of tutors consider plagiarism from internet as the major problem. In current teaching-learning systems, the identification of problem situations in student learning processes, are largely performed after they actually happen. This approach is however ineffective in addressing the needs of future learning environments.

A learning system architecture called ICARE [1] [4] has been proposed to overcome the problems mentioned and to serve as a model for future e-learning systems. Details of the ICARE concept can be found in [5]. Key ideas of this ecosystem are summarized as follows:

- “Support for students to be actively engaged in creating their own intellectual property and respect other people’s copyright.
- Provide a guided environment to foster constructive learning practices and critical thinking whereby students can structure their reading and construct written works systematically.
- Support for the acquisition of basic reading-writing skills i.e. paraphrasing, summarizing and referencing accurately.
- Support for process management to structure course-work as a series of process steps. Student learning can then be continuously checked and assessed at each stage.
- Incorporate preventive measures for making sure that unwanted versions of copy-and-paste just cannot happen (or is drastically minimized)
- Incorporate viable technologies required to minimize the supervision effort required from instructors.” [1]

As plagiarism and copy-paste activities are not restricted to text documents, a more comprehensive addressing is required. This paper then further describes the architecture and design of an integrated multimodal plagiarism and similarity detection system. Layered similarity assessment services incorporated in described system not only provide means of semantic aware plagiarism detection, it also guides students during the learning phase. ICARE in our work is introduced as a platform of hybrid learning and similarity assessment services.

II. Learning Approach In ICARE

In contrast to traditional e-learning, the adapted learner centered approach supports the more fluid form of e-learning with dynamic acquisition of knowledge and skills. Multiple learning scenarios are created and presented to user where learners can assess the situations and react appropriately. The blended learning approach of ICARE allows the stakeholders to share learning experience. Learner's evaluation and outcome of learning activity is based on learner's interaction with system rather than mere evaluation of end result. Such learning environments help in building indicators of successful learning outcomes. The ICARE ecosystem is being built on top of WBT-Master [6] a sophisticated e-learning system that supports the definition of multiple learning scenarios, project-based administration of e-learning and the interactive classroom management activities such as mentoring, brainstorming, project management etc. [7]. An overview of the ICARE ecosystem is shown in Figure 1.

A rigorous academic reading and writing process is thus enforced in a guided environment. Each student maintains a portfolio, which characterizes and represents all learning outcomes, recognized student works and achievements. An e-diary is employed to enable the aggregation of student (individual and collaborative) contributions to be captured, assessed and reflected upon. Internal processes and states of the learners can then be represented and augmented with systemic input to provide deeper insights on their learning.

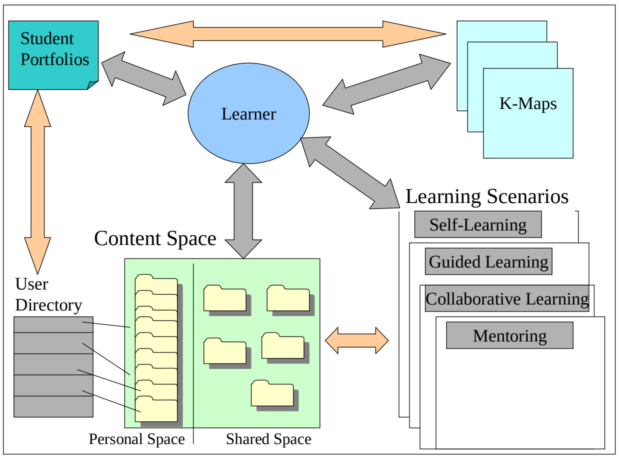

Figure 1. Overview of ICARE Ecosystem

The ecosystem allows the demonstration of student learning and understanding via knowledge maps (K-Maps). K-Maps can be applied at an appropriate stage in the learning process, enabling instructors to determine the actual learning of students to make an informed judgment on possible cheating on the part of students. This proposed system also allows the instructor or mentor to identify areas in which students found difficulty in understanding or assimilation. The Copy-Paste checking subsystem has been integrated directly into the e-learning environment in such a way that it directly monitors student activities and provides appropriate feedback to instructors, mentors and students themselves. Multiple modes representing alternative sources of feedback on student learning is consolidated to provide insights into student learning. We will describe the overall system architecture of ICARE before focusing on the similarity checking facility.

III. Mashup based Architecture

Mashup [8] application architecture is one the most important offshoot of web 2.0 phenomenon. It is an emerging composite web application development paradigm where applications are constructed from the services and content from other web sites. Mashup introduces a lightweight and dynamic model for combining contents and functionalities. It enables the rapid development of flexible applications built upon a collection of Web Services. Web services provide interoperable machine to machine interaction over the World Wide Web. The standardized xml based messaging, service description and discovery enables the convenient and widespread use of this specific form of service oriented architecture. Mashup architecture has been proposed for ICARE, built upon a set of Web services as shown in Figure 2.

There are a number of services, which are required to facilitate learning management in ICARE. These services can be separated into basic services such as K-Card management, profile management, session management, etc. and application specific services which includes concept discovery management, expertise finding, knowledge visualization, Copy-Paste administration (with broader coverage of textual and visual contents), etc. Current works at Graz University of Technology include the development of a number of these Web services, as described in [9], [10], [11], [12].

A number of databases are then employed in the realization of the e-learning ecosystem to support learning management, student activity tracking and knowledge visualization.

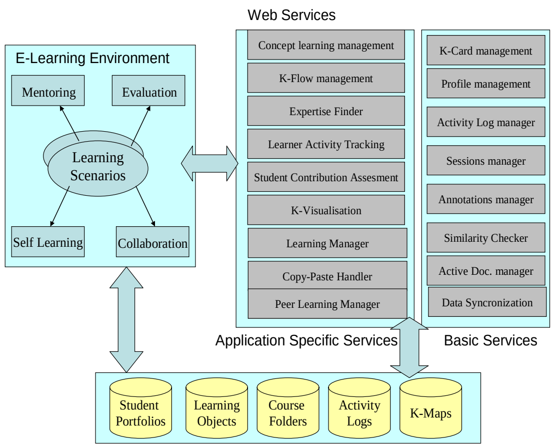

Figure 2. Mashup based Architecture of ICARE

We will now focus on the layering of the Web services to support the integrated semantic aware similarity checking facility. Figure 3 describes the layering applied in the design of a sophisticated mashup with a focus on the checking of Copy-Paste activities.

Data access and checking service is performed in the background, while students are engaged in their learning activities. Access to both internal databases and the Web is required in finding for relevant material related to a particular student work. In order to access web resources a number of internet search engine APIs are used. The course content space and learner's profile is also accessed in order to build a rich knowledge base. In the context of Copy-Paste checking, this involves finding for documents that may be copied from or is similar in concept to the student's current work.

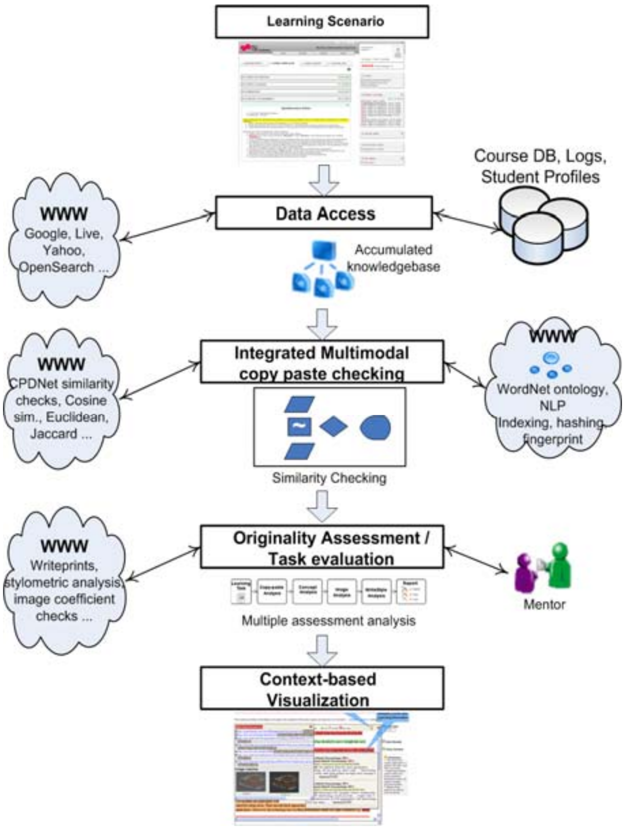

Figure 3. Layered Web services Architecture

An integration service is then performed to consolidate the data fragments from multiple sources. An ontology is referred to in this respect to identify information that is most relevant to a student's task or context of work. A number of similarity checking services are used to determine links of matching contents. The language ontology is also required to check for media-independent semantic-level conceptual document similarity. Once information that was considered to be relevant has been pushed to the user in a non-obtrusive manner, the system continuously analyses the feedback of users (students, peers, mentors, etc.) to determine the value of the information supply. Based on a deeper analysis, students are advised and supported appropriately. The results of the analysis are then visualized and presented to the student.

As opposed to current plagiarism detection methods as used in educational institutions, [13], [14], a cross-media conceptual similarity checking approach [15], [16] has been adopted. We will now review the mechanism for integrating multimodal Copy-Paste checking and design of the similarity checking services.

IV. Mechanism for Integrating Similarity Assessment Services

Activities in ICARE's learning scenario can be tracked through the events happening in student workspace. The student's workspace is a document processing environment in ICARE. Instructor can choose to monitor this editing environment in order to track progress of the learning task. Any changes in workspace are analyzed through selected content analysis tools. The student is facilitated by a proactive information supply during the learning and writing phase. This feedback mechanism provides the links to matching resources without conventional active search. Figure 4 illustrates the integration and use of similarity assessment services in ICARE learning environment.

The similarity assessment manager is activated by an instructor to perform originality assessment on a selected set of archived documents or to evaluate a student workspace. The assessment manager calls a number of selected web services making use of a selected set of content analysis techniques. The activated similarity checking services find the matching contents found in local content space and over the internet. The source is evaluated for intrinsic write-up consistencies and stylometric [17] comparison is made with archived work in student's profile. A layered similarity detection approach is adapted in ICARE in an effort to provide a more semantic aware plagiarism detection system. Details of these techniques are described in following section. Information aggregator compiles the results collected from various services and enriches the student workspace with related resources. It is also responsible for preparing an originality assessment report for the instructor and mentor.

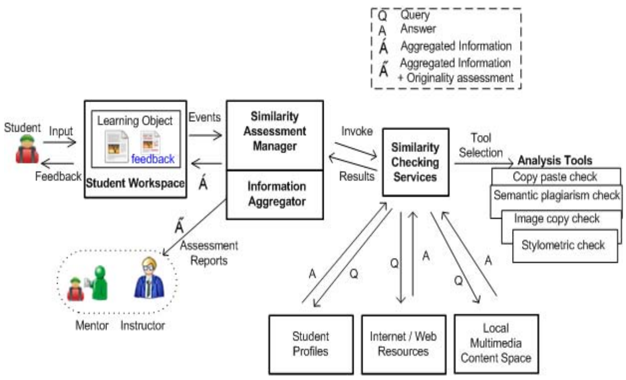

Figure 4. Design of Integrated Plagiarism Checking

In order to materialize the idea of ICARE eco-system the WBT-Master is being augmented by a number of similarity assessment services from CPDNet [18]. This allows a convenient integration and extended use of plagiarism detection process. Similarity detection is introduced during the learning activity to provide a better understanding of task at hand to support the student learning process.

The interface of the prototype application is shown in Figure 5; the information supply feedback in the student's work space highlights matching text and visual contents. This information improves student's understanding of intellectual property rights and helps prevent potential infringements. The key component that supports information supply is the multimodal similarity detection services. The similarity detection capability has been provided as web services. We will now explore how these services are further applied to perform similarity detection operations beyond simple copy-paste detection. Feedback on student learning activities is further augmented by stylometric information as evidence of originality. A consolidated measure of similarity detection is computed based on information made available from multiple sources of information: writing style similarity, similarity within text documents and similarity based on image analysis. This approach promises a more comprehensive approach in learning pattern discovery (also addressing copy-paste detection). Our current experiments explore the cross media copy-paste checking by matching media annotations with normalized terms in text.

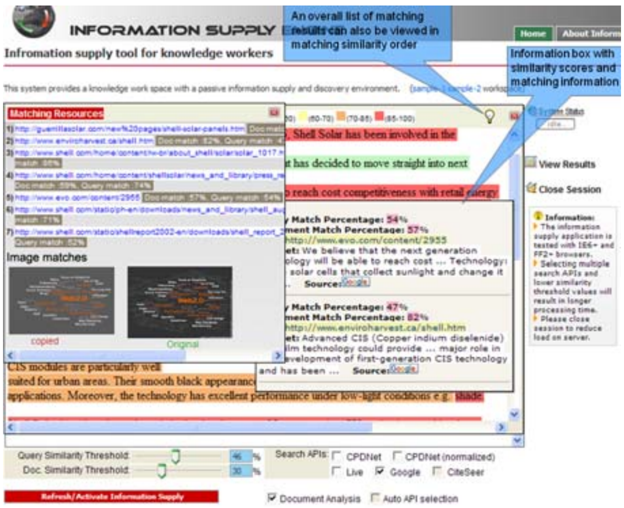

Figure 5. Information Supply System [19].

V. Similarity Detection Services: Beyond Text Copy-Paste

The measurement of similarities between information objects is the key function of any information retrieval, and knowledge discovery application. Factors like exponential growth of digitized data, increased information access due to enhancements in communication links and devices, expanding and collaborative nature of internet applications; demands the discovery of a deeper level of relationships among information entities. However, current systems do not go beyond a standard term matching similarity detection. Furthermore, learning systems makes use of multiple media types and sources, contents are authored and composed by a number of persons in varying environments. In situations like this, simple copy-paste detection using word match is not enough to determine consistent patterns and deeper relationships. ICARE learning environment aims to provide services that goes beyond the conventional copy-paste checking mechanism. This is done by adding a number of similarity checking services. These services are available as part of broader plagiarism detection environment CPDNet. The following subsections describe these services.

A. Term semantics aware text similarity checking service:

Text similarity is a basic function that determines the degree of similarity between a document to be evaluated and document(s) found either on the Web or in an internal student database. In common the plagiarism detection systems used by academia compare exact words for finding documents with similar text. However as reported by a survey of these tools [15], this approach tends to fails when paraphrasing is introduced. Simple experiments revealed that using dictionary assisted tools one can systematically replace words in text with synonyms, rendering conventional plagiarism detection systems useless. In traditional learning environments the copy-paste plagiarism checks are initiated after the submission process. There is also no mechanism to monitor and control the duplication of text during the learning phase. ICARE address the latter issue by adapting an approach where an instructor can verify student inputs at any point in their learning activity. The student work space is actively monitored in the background; instructor can request a service to compute workspace similarity scores with internal and external learning content space at the different stages of leaning task. The student's knowledge workspace is linked with a number of search services which provide up-to-date information about the task at hand without performing a conventional active search. This information not only determines the ability of students in carrying out a task, but also if given access to intermediate feedback student can learn the rightful means of performing an academic task [10].

The service used by ICARE provides enhanced conceptual text similarity detection in order to address the earlier mentioned issue of copying paraphrased form of similar text. We will describe added semantic similarity detection in more detail after describing the general steps taken to discover and compare matching documents. This service takes as input a text that needs to be checked for similarity with either an existing document in the internal collection or the Web or from the active workspace of student.

The input text is first broken down into moderately sized segments called fingerprints. The target index is identified from local and internet collection, this index is used for document similarity checking. The access and selection to the checking repository is made through the local and web searching APIs. The generated fingerprints are then used as search queries to identify a list of suspected similar document from available search services. The size of snippets of text used as fingerprints is twenty words, which is approximately equal to the size of snippets returned by general web search engines such as Google, Live and Yahoo. Snippet sizes can however be varied for extracting either coarse-grained or fine-grained source text segments as required. Searching APIs at server side uses matching algorithms (in common distance measures applied on weighted term vectors) and proprietary ranking algorithms to produce matching document lists. Equation (i) shows the similarity measure used by system's local searching API based on Lucene[1].

$$\text{score}(q,d) = \text{coord}(q,d) \cdot \text{queryNorm}(q) \cdot \sum_{t \text{ in } q} \left(\text{tf}(t \text{ in } d) \cdot \text{idf}(t)^2 \cdot \text{t.getBoost}() \cdot \text{norm}(t,d) \right) \quad \text{(i)}$$

An added layer of relevance detection is introduced that examines the term match between fingerprints and snippets of returned results. This allows us to eliminate the page ranking noise added by multiple search engine APIs. The added similarity check allows a combined and seamless use of common search services for plagiarism specific checks. A simpler similarity measure, shown in equation (ii), based on inclusion exclusion principal is sufficient to perform the query and snippet level comparison.

$$\text{Jaccard similarity}(q,s) = n(q \cap s) / n(q \cup s) \quad \text{(ii)}$$

A report is then presented to highlight the matched fingerprints and the corresponding degree of copy. Document similarity is then computed based on the extent of matching fingerprints in a target document. A primary suspected document list is first constructed based on matching fingerprints. Further similarity checking is then performed on the primary list at the document level taking into consideration the number and order of matching fingerprints within a document.

The described process flow of similarity checking service plagiarism is illustrated in Figure 6. It shows compound use of a set of web services to generate desired results.

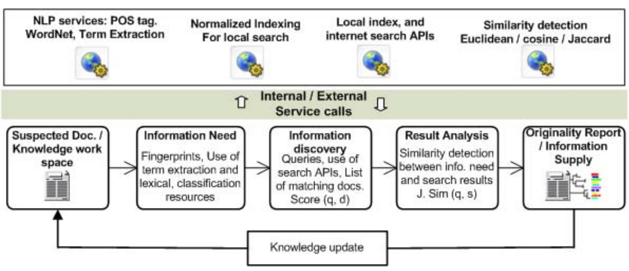

Figure 6. Flow of document similarity check in ICARE

[1] http://lucene.apache.org/

As mentioned earlier traditional plagiarism techniques employ word based or text hash based similarity checking, this approach fails when a systematic variation in text is introduced. The text normalization approach part of ICARE plagiarism check service is an effort to minimize this effect. A generic or canonical form of words in text is taken into account while performing similarity checks. WordNet [20] ontology is used as a means to enable a deeper concept level checking. Figure 7 illustrates the text normalization procedure for a better understanding. The local content space of ICARE is indexed using CPDNet's indexing engine. The text of content space is extracted from various media types (PDF, MSWord, RTF, HTML, XML, PPT) and part of speech tagger is used to determine the lexical category of terms in text. In general information retrieval applications ignore stop words and little importance is given to the function words (words with little lexical meaning) while calculating relevancy scores. The major content words (verbs, nouns, adjectives) carry most of the weight while computing similarity among text. The content words indentified in text are looked up in the language ontology with a matching grammatical sense. This word sense disambiguation helps to identify a related set of synonyms in language ontology. The most significant form of word in determined synonym group is further determined through use tag_count property of WordNet which represents the reference counts in language. Transforming the contents words in that significant form creates a generic representation of text. The text being compared for similarity thus provides a deeper level of alternate term match.

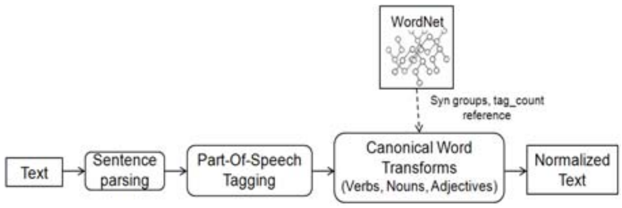

Figure 7. Conceptual text similarity checking approach

The local content space is stored as an inverted file index and is available to ICARE search. The word vectors of queries and index are compared for similarity using dot product or cosine similarity criteria. The normalized form of texts allows checks based on normalized root concepts in addition to the typical word based matching. The text plagiarism detection design of ICARE as shown in Figure 6 is a distributed application. It shows composite use of web services to compare both the internet and locally shared document sources. Service Oriented Architecture (SOA) is employed to support and connect a set of heterogeneous applications. Collaborative units (web services) work transparently within a single computing environment. ICARE exemplifies the use of standardized, platform-independent web enabled components to build complex computing environments. It shows signifies a shift of information systems towards scalable information services.

B. Similarity checking of multimedia contents

Another missing link in plagiarism detection application highlighted by different reports [15][21] is the ability to check for similarity beyond text. Multimedia contents, especially images are an important part of learning material or documents. E-learning contents, research papers and student assignments contains far more than words, they contain a mixture of pictures, drawings, maps, graphs equations and symbols. These elements are beyond the relevance detection scope of a common search and indexing application. In commonly available learning environments and almost all plagiarism detection systems Content Based Retrieval systems (CBR) has not been exploited effectively. In ICARE we introduce the use of Content Based Image Retrieval (CBIR) services to add similarity detection of visual contents. For that matter images available in learning content space are processed for feature extraction. The process of feature extraction involves application of number image analysis technique through which the input data (image) is transformed into a reduced representation of feature set. The analysis techniques include Fourier Transforms, Wavelet Transform, Gabor Transform, Hough Transform, Edge, Blob, Corner detection algorithms [22] etc. The selected visual features of images in documents are stored in suitable indexes along with text indexes. When an instructor launches the similarity assessment services for images from student's work space, target document are acquired and processed for visual feature extraction. A search query is launched for identifying images in learning content space having similar visual characteristics. A number of open source CBIR [23][24][25][26] systems are available that can be incorporated in similarity assessment process of ICARE.

Major strengths of a CBIR include the following:

- Efficient extraction of the visual characteristics from image.
- Creation, maintenance and efficient storage (index) of image describing characteristics.
- Provision of user friendly query mechanism (e.g. query by example, query by selection from index etc.)
- Effective result presentation: access to ranked results, and similarity scores etc.
- Possibilities effecting ranking (added weight to specific characteristic matching; color, texture, shape)
- Feedback mechanism
- Ease of integration: Programmatic access to the CBIR functionalities

We evaluated three open source CBIR systems; table 1 gives a comparative overview of their features.

TABLE I. COMPARISON OF CBIR FEATURES

	GIFT[24]	*LIRE[26]*	*isk-daemon [25]*
Visual Feature Space	Color histograms, color layouts, possibilities of adjusting feature space for a number of color and grey scale values	Feature space based on MPEG-7 descriptors, color histograms, scalable colors, color layout, edge histograms.	Image signature using multi resolution wavelet decomposition, gives reasonable image approximation with smaller feature space, support for video frame feature extraction
Indexing	Inverted file index for feature storage, indexing in single go, no flexible API for indexing update and maintenance.	Lucene based inverted file index, for efficient storage and access. Ease of update maintenance due to availability of tools, programs to manage Lucene index	Internal binary format for image signature storage, Web admin interface for index management.
Query	Query by example, random image selection from index	API gives example of development of various query scenarios.	XML-RPC interface for browsing index, and possibilities of creating client to upload image in index and perform query by example.
Results	Allows use of multiple weighting algorithms for ranking, e.g. classical IDF, separate normalization, User feedback for query refinement	Ability of selecting manual weights for visual characteristics for application specific filtering	Ranked output with fixed internal relevance scoring algorithm.
Ease of integration and extended use	Standardized MRML interface, example client codes, web interface	Modular Java based API, with convenient integration and extension to code.	Python based API, with code examples for linking web clients to server, distributed operation for scalability

CPDNet platform responsible for providing similarity assessment services to ICARE uses LIRE image analysis and retrieval API. It is selected due to its compatibility with the text based indexing and similarity assessment services used in our system, which also use Lucene compliant indexing. The content indexing service, while generating document index, extracts the images from documents and stores the visual features in extended inverted file index.

An added layer of originality assessment is introduced in learning environment that determines the originality of queried image and matching images in content space. In a case where a student adds an image in its document and CBIR identifies a very similar image in available content space. The added layer of image originality assessment determines which one of the similar identified images is the original. Meta information, such as and content-centric attributes can be taken the criteria to determine originality. In cases where limited meta information is available the image quality assessment criterion is then used to determine the originality in similar set of images. A method used for blind images quality assessment determines the blurring artifact introduced in copied images through DCT/Wavelet coefficient analysis [27]. The generally adapted lossy compression algorithms remove the higher frequency information in image. This effect can be determined by analysis of quantized coefficients. Future versions of our prototype will incorporate a semi-automated facility for acquiring user feedback in labeling of images. Operation of similarity assessment for visual contents in ICARE is described in figure 8.

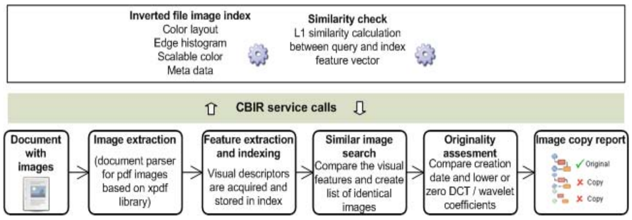

Figure 8. Flow of image similarity check in ICARE

VI. WRITEPRINT ANALYSIS

Intrinsic characteristic of text are often ignored by computer applications while checking for plagiarism in documents. These characteristics include stylometric elements such as structure of document, sentence lengths, use of words, use of special phrases, use of function and stop words, use of punctuations etc. Research in this area indicates that these features can be very effectively used to identify the authorship of documents [28], or highlight inconsistencies in text. ICARE learning system allows comparative analysis of stylistic elements of text. The services acquired from CPDNet partner project generates writeprints using sentence lengths, word sizes, and writing style feature matrix based on the use of punctuations, articles, pronouns, conjunctions. Figure 9 shows a sample writeprint analysis application developed at TUGraz. It demonstrates the use of writing characteristics to identify similar contents with common authors.

Figure 9. Write print analysis for text similarity assessment

VII. CONCLUSION

The next generation of e-learning environments will not be packaged as rigid systems. The current state of e-learning and unchecked student expressions have called for the focusing on learners with just-in-time learning support. This paper described the layered architecture for ICARE, with details of its realization as a mashup. It has further described the design aspects of the integration of a multimodal Copy-Paste checking facility into an E-learning system. The Web Services based design has demonstrated means of overcoming the limitations of current e-learning systems, which delay the checking of infringements until students submit their assignments. The modular approach applied in the design of ICARE allows these novel ideas to become adapted to other e-learning systems beyond WBT-Master. The specification of the overall architecture and design also serves as a basis for the design considerations of future e-learning systems.

REFERENCES

[1] N. Kulathuramaiyer, B. Zaka, D. Helic, “Integrated Copy-Paste Checking: Design and Services, Advanced Learning Technologies, ICALT '08. Eighth IEEE International Conference on 1-5 July 2008 Page(s):519 - 523 Digital Object Identifier 10.1109/ICALT.2008.312
[2] D. McCabe, “Assessment Project Research survey” Center for Academic Integrity, Retrieved Nov. 5, 2008, from http://www.academicintegrity.org/cai_research/index.php
[3] ATL survey, Retrieved Jan. 25, 2009, from http://www.atl.org.uk/media-office/media-archive/School-work-plagued-by-plagiarism-ATL-survey.asp
[4] N. Kulathuramaiyer, H. Maurer, “Coping with the Copy-Paste Syndrome”, Proceedings of E-Learn 2007, AACE, Calgary, Canada, 2007.
[5] N. Kulathuramaiyer, H. Maurer, “Addressing Copy-Paste with ICARE”, accepted for publication in National University Journal, USA, June 2008.
[6] “WBT Master White Paper”, Retrieved from http://coronet.iicm.tugraz.at on March 25, 2008.
[7] D. Helic, H. Krottmaier, H. Maurer, N. Scerbakov, “Implementing Project-Based Learning in WBT Systems”, Proceedings of E-Learn 2003, AACE, USA, 2003, pp. 2189-2196.
[8] N. Kulathuramaiyer, “Mashups: Emerging Application Development Paradigm for a Digital Journal”, JUCs, Vol. 13 No. 4, 2007, pp. 531-543. http://www.jucs.org/jucs_13_4/mashups_emerging_application_development.
[9] M.T. Afzal, N. Kulathuramaiyer, H. Maurer, “Creating Links into the Future”, Journal of Universal Computer Science, Vol. 13, No. 9, 2007 pp. 1234-1245. http://www.jucs.org/jucs_13_9/creating_links_into_the.
[10] B. Zaka, H. Maurer, “Service Oriented Information Supply Model for Knowledge Workers”, Proceedings of I-Know 07, 2007, pp. 432-439.
[11] S. Khan, N. Kulathuramaiyer, H. Maurer, “Applications of Mashups for a Digital Journal” accepted for publication in Journal of Universal Computer Science.
[12] S.C. Ong, N. Kulathuramaiyer, A.W. Yeo, “Automated Discovery of Concepts from Text”, Proceedings of the IEEE/ACM/WIC Conference on Web Intelligence, 2006, pp. 1046-1049.

[13] Turnitin®, Retrieved from http://www.turnitin.com on December 9, 2007.
[14] Mydropbox®, Retrieved from http://www.mydropbox.com on December 9, 2007.
[15] H. Maurer, F. Kappe, B. Zaka, Plagiarism: a Survey, Journal of Universal Computer Science, Vol. 12, No. 8, pp. 1050-1084, 2006.
[16] N. Kulathuramaiyer, H. Maurer, "Fighting Plagiarism and IPR Violation: Why is it so Important?", Learned Publishing, Vol. 20. No. 4, pp. 13-19, 2007.
[17] S. M. Eissen and B. Stein, "Intrinsic Plagiarism Detection", LNCS, Volume 3936/2006, Advances in Information Retrieval, Pp 565-569, DOI: 10.1007/11735106
[18] CPDNet (Collaborative Plagiarism Detection Network) Retrieved from http://www.cpdnet.org on Jan. 05, 2009
[19] "Information Supply Engine", Retrieved from http://www.cpdnet.org/isdemo on March 25, 2008.
[20] G. Miller, "WordNet : An Electronic Lexical Database", MIT Press, Cambridge, MA. http://wordnet.princeton.edu/
[21] G. Chester, "Final Report on the JISC Electronic Plagiarism Detection Project" Joint Information Systems Committee, August 2001, Retrieved Nov. 20, 2008, from http://www.jisc.ac.uk/uploaded_documents/plagiarism_final.pdf
[22] M. Sonka, V. Hlavac, R. Boyle "Image Processing, Analysis, and Machine Vision", 3rd edition 2007 ISBN: 049508252X , 9780495082521
[23] RC Veltkamp, M. Tanase, Content-based image retrieval systems: A survey. Technical report UU-CS-2000–34, Department of Computing Science, Utrecht University, October 2000
[24] GIFT (the GNU Image-Finding Tool) Retrieved Jan. 05, 2009, from http://www.gnu.org/software/gift/
[25] imgSeek Content Based Search, Retrieved Jan. 05, 2008, from http://www.imgseek.net/
[26] LIRE (Lucene Image REtrieval) Retrieved Jan. 05, 2009, from http://www.semanticmetadata.net/lire/
[27] B. Zaka, M. E. Steurer, F. Kappe, "Framework for extending plagiarism detection in virtual worlds" submitted for publication in RCIS 2009.
[28] A. Abbasi, and H. Chen, Visualizing authorship for identification. In Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics, San Diego, CA, 60–71.

Narayanan Kulathuramaiyer is a Professor of Computer Science at the Faculty of Computer Science, University Malaysia Sarawak. His current areas of interest include Semantics-Aware systems, Assimilative Intelligence and Technology-based learning systems. He has authored over 60 papers in peer reviewed journals and conferences.

Bilal Zaka holds Ph.D. degree in Computer Science from Graz University of Technology, Austria. He has previously served as a Systems Manager at COMSATS Institute of Information Technology, Pakistan. At present, he is a research fellow at the Institute for Information Systems and Computer Media, Graz University of Technology Austria. His research interests include collaborative web applications, knowledge management, multimodal applications for the internet and computer networks.