

A High Performance Optoelectronic Machine for Automated Arabic-English Parallel Corpus Creation and for Text Mining Processing

Samy S. A. Ghoniemy and Omar H. Karam

Abstract—In this paper a parallel optoelectronic computer architecture is proposed for large-scale parallel corpus, full text search and text mining applications while achieving high speed and high performance and utilizing the parallel processing nature of optics. An optical technique is presented that aims to automatically construct an Arabic-English corpus through web mining. In addition, it is shown how Arabic morphology and Arabic synonymous can be used to extract the required knowledge from Arabic texts based on converting the Arabic text into a semi-structured database. The proposed system is modeled and implemented using a two-dimensional space optics symbolic correlator to perform the comparison operations of the expert system. The proposed optical engine is able to perform 3D processing for inferring conclusions to queries by representing page contents in a 2D space using mapped-template logic. A hardware realization for these systems based on commercially available components is discussed. The performance analysis of the proposed system shows its remarkably improved processing performance with respect to processing speeds and both recall and precision measures.

Index Terms—Optical computing, parallel corpus, text mining.

I. INTRODUCTION

With silicon based technologies for logic already approaching the physical limits of switching speed [1], [2], Optics may have answers for these limitations. In addition, due to continuing globalization there is also much interest in multi-language collections. The recent availability of machine translation systems is, in that context, an important development. Multi-language text mining is much more complex than it appears. In addition to differences in character sets and words, text mining makes intensive use of statistics as well as the linguistic properties (such as conjugation, grammar, senses or meanings) of a language. There are many basic assumptions about capitalization and tokenization that work for one language and that would not work for other languages. When text mining techniques are used on non-English data collections additional challenges have to be addressed. All these reasons are require the development of new computer techniques/systems to help

Manuscript received May 15, 2013; revised July 19, 2013.

S. Ghoniemy is with the Faculty of Informatics and Computer Science, The British University in Egypt and MTC, Cairo, Egypt (e-mail: ghoniemy@cis.asu.edu.eg).

Omar H. Karam is with the Faculty of Informatics and Computer Science, The British University in Egypt and the Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt (e-mail: omar.karam@bue.edu.eg).

control and harvest huge amounts of information.

Inherent features of optical components make them ideal for transferring planes comprising millions of data points simultaneously and thus achieving effective multi-dimensional input/output and parallel processing at the same time. [3]-[12]. The optical bi-stabilities recently developed with switching times in the order of Femtoseconds [13], FS, compared with milliseconds in electronics, have increased the possibility of using optics for the practical realization of new parallel optical computers for optical information processing especially in the field of Knowledge base systems (KBSs) and text mining as they have reached the Terabyte level [14]. The text mining involves processes such as information retrieval, text analysis, information extraction, clustering, categorization, visualization, and machine learning [15]. These require powerful processing methods and involve extensive parallel processing.

Cross-language information retrieval (CLIR) is another highly required application for retrieving information written in a language different from the language of the user's query. 43.4% of the web content is in English while the English web population does not exceed 26.8% [16]. Therefore, CLIR has many useful applications such as searchers with a limited active vocabulary, but good reading comprehension in a second language (e.g. English), might prefer to issue queries in their most fluent language [17]. A large-scale parallel corpus plays an important role in CLIR by providing the primary training data for the statistical translational models [18]. Most of the previous work on parallel texts has been conducted on a few manually constructed parallel corpora such as the work published by Canadian Hansard Corpus and Linguistic Data Consortium (LDC) [19].

The main problems in building parallel corpora are accuracy and processing time. Since such systems are always real time systems they still require higher processing speed. Implementation of parallel unification for high-speed matching is currently under investigation [14]. Herein lies the potential of an optical implementation of optical text mining machines that are compatible with electronics to exploit the strengths of both.

The main objectives of the system proposed in this paper are: i) constructing the optically-based parallel corpus (OPC) automatically through web mining, and ii) preparing the developed OPC to be used as training material for the translation model. The proposed system will be used for English-Arabic parallel texts to construct the corpus and can be extended to include other language pairs in a similar way.

The remainder of the paper is organized as follows: In Section II, the proposed optical parallel corpus is presented.

The architectural design of the overall text mining and parallel corpus machine and the optoelectronic design of each subsystem are discussed in Section III. In Section IV, the hardware realization of the proposed system is illustrated. Section V presents the experimental results and the performance evaluation of the proposed system. Finally, the summary and conclusions are in Section VI.

II. PROPOSED OPTICAL PARALLEL CORPUS

The proposed optical mining system functionalities and operation sequence are illustrated in Fig. 1. The proposed English-Arabic parallel corpus system is intended to be an online real time system which starts after the user issues its query in the form of an http request (URL on a specified host). Issuing this request starts the optical host crawling algorithm which crawls the host for all pages that are written in either Arabic or English by executing a language detector algorithm which identifies a list of parallel language-pair page patterns and stores them in a structured file and filters out all other pages in the domain as shown in Fig. 2. The output from this algorithm is a set of candidate pairs. The next step is the execution of the contents optical extractor which filters out the HTML tags and other scripts and keeps only the clean page contents. Finally, content-based matching is performed to calculate the parallelism similarity between each candidate pair using an English-Arabic dictionary. The output of this system is an English-Arabic parallel corpus that is well-aligned at paragraph level with completely clean texts. This approach is straight forward, fully automated, and easy to port to any other language pair.

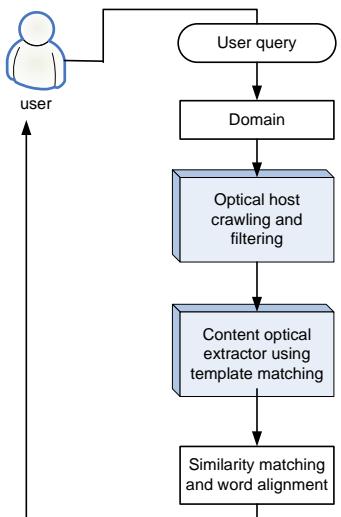


Fig. 1. The proposed optical mining system.

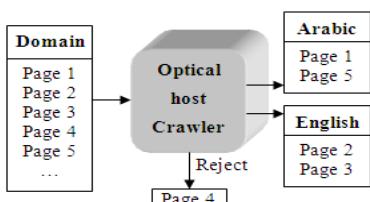


Fig. 2. The proposed optical host crawling process

III. DESIGN OF TEXT MINING AND PARALLEL CORPUS MACHINE

The proposed hybrid optoelectronic architecture is shown in Fig. 3. It is capable of processing data and performing text searches optically and in real time. In this architecture, all the Web documents are considered to be stored in banks of large optical disks. The output light is then spatially modified using optical elements that form the Pattern Generation Unit (PGU). When a user issues a request (query), the participating Web pages in the specified domain are retrieved and the PGU then forms an appropriate beam distribution. The optical data are then processed by the optical data processor and the result is projected on a photodetector array.

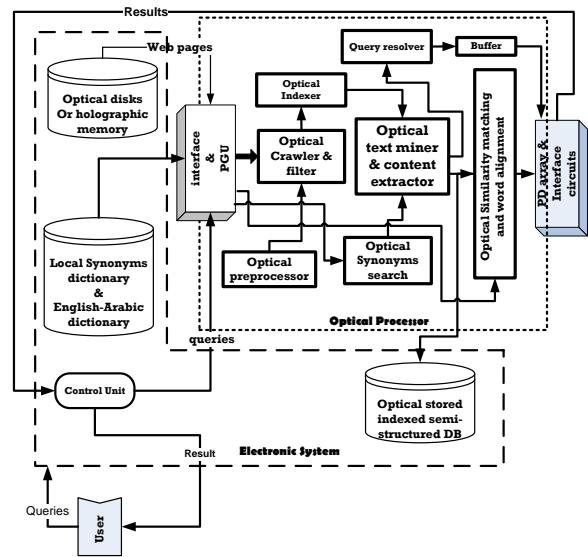


Fig. 3. Proposed optoelectronic architecture for the parallel corpus machine.

The proposed optical machine consists of three main subsystems:

- 1) An electrical-to-optical transducer to convert electrical signals to 2-D optical signals.
- 2) An optical logic device (optical processor) to perform the encoding, crawling, indexing and computations in 2-D where all operations are optical.
- 3) An optical-to-electrical transducer to transform optical signals into electrical signals.

A. Web Pages Encoding

In this paper, the technique adopted is storing all the domain downloaded pages in a page oriented holographic memory (POHM) that can be efficiently realized and processed in parallel form. In the proposed system, both optical data and queries are encoded using optical symbolic substitution (coding rules) [20], [21] and Web page contents are encoded in binary matrices (patterns). The search algorithm is implemented using a set of logical operations in parallel. With a relationship matrix, the optical host crawler process is very similar to an inner product operation like matrix-vector multiplication. The optimum design of this machine effectively combines the individual strengths of optics and electronics. In such an “optoelectronic” (OE) system, electronic elements are used to control the optics and to provide the input and output interfacing devices of the optical elements or subsystems.

B. Relationship Matrices for Indexer Implementation

Since relationship matrices are processed via a matrix-vector product, a learning process can be realized simply using rules operation (matrix-matrix multiplication) and implemented optically using the above proposed optical correlator as shown in Fig. 4 [22], [23].

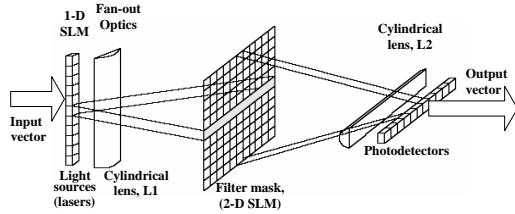


Fig. 4. $N \times N$ parallel matrix-vector inner product processor.

In this figure, an input light beams representing a set of symbols is applied to the system using a 1-D SLM. Each light beam is spread horizontally using the cylindrical lens L1 across the filter mask whose transmittances are related to the values in the relationship matrix. This mask can be implemented using a 2-D SLM. The light is then focused using the cylindrical lens L2 into a linear photodetector array that produces the output vector which represents the result of the matrix-vector inner product. Using the same methodology we can implement the matrix-matrix inner product if we used 2-D SLMs in the input and output instead of 1-D SLMs.

C. Optoelectronic Comparison Unit for Web Pages Matching and Similarity Measure

In this section we adopted the work published by Guilfoyle [23] in which optical comparisons can be performed based on the exclusive-or (EX-OR) primitive using dual-rail logic. Since both the value of the bit and its complement are needed for the comparison, each n-bit word will be represented by $2n$ light beams (i.e. the 4-bit word 1101 will become 10-10-01-10). Using this method, a 00 combination corresponds to a “don’t-care” character while the 11 combination always produces a “Not-Equal” result. The coding scheme for the two logical values, 1 and 0, can be light and no light respectively. The proposed optoelectronic comparison unit is shown in Fig. 5.

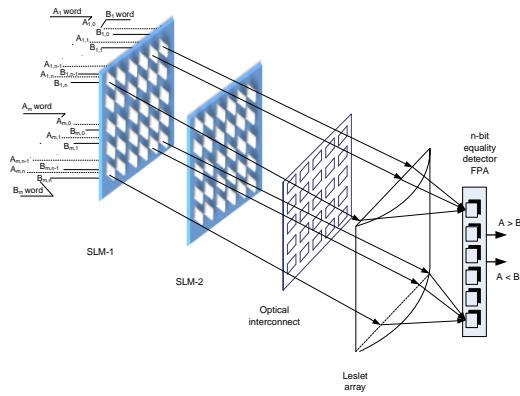


Fig. 5. An optical comparator of m n -bit words using XOR rule

Fig. 8 shows that the light beams are superimposed bit-wise and are focused using a cylindrical lens on a single

photodetector which performs the logical OR (or summation) of all the beams [22]-[24]. If no light is detected the two words are equal while any level of light intensity other than zero indicates that the two words differ in at least one bit. Multiple word comparisons can be performed in parallel if two 2-dimensional arrays, A and B, are employed each having $2n$ rows and m columns. At any instance in time the word at the i^{th} column of the A array is compared to the word at the i^{th} column of the B array. The result (equal or not equal) is recorded on the i^{th} cell of a row of photodetectors. Figure 6 illustrates how m comparisons of n -bit words take place simultaneously.

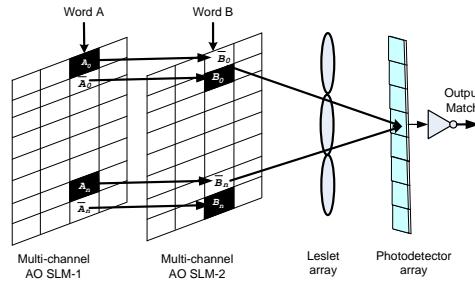


Fig. 6. Encoded two array optical comparator for m n -bit words comparison.

The information on each array is recorded on 2-D multi-channel Acousto-optical (AO) spatial light modulators (SLMs). These active optical devices store encoded input pattern, and/or spatially modify or amplify some of the optical characteristics (phase, amplitude, intensity, polarization) of a readout light distribution as a function of space and time. In this paper we have used a reflective optically addressed SLM-1 (LC-R 1080, HOLOEYE) and a transmissive electrically addressed SLM-2 (LC 2002, HOLOEYE) [25].

IV. HARDWARE REALIZATION OF TEXT MINING AND PARALLEL CORPUS MACHINE

To implement a relationship matrix form of the Web pages contents, the optical mapping templates approach is adopted. These templates are binary transmission masks (in which transparent squares represent logic 1's and the logic 0's are represented as opaque background squares). The first few character positions on SLM-2 will permanently contain the character strings for *New-Paragraph*, *New-Sentence* and *New-Word* (we used the indicator “■” to refer to the *New-Word* character). The remaining length will hold the search argument(s). In Fig. 7, white squares represent logic “1’s”, while others represent logic “0’s”.

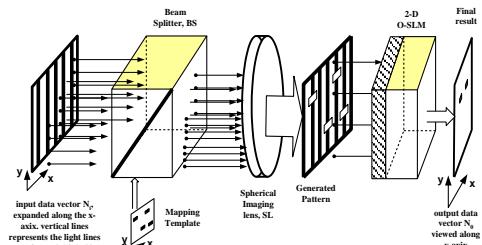


Fig. 7. A conceptual realization of the mapping templates

In this figure an input vector N_i , is vertically expanded

along the y-axis (to form an array, each row of which equals N_i), and then optically overlaid with the mapping template using imaging optics (SL), then a 2-D OSLM performs a 2-D logical AND operation. The resulting output N_o is viewed along the rows of the output 2-D photodetector array which gives the query solution.

Using this conceptual design of the optical processor that is presented in Section III, a complete optical processing machine including optical crawler, optical indexer, optical text mining unit, and mapped-template logic shown in Fig. 10 is shown in Fig. 8. The hardware design of the proposed optical processor consists of 2 E-SLMs, 2 O-SLMs, 4 spherical imaging lenses (L1, L2, L3, L4), 5 beam splitters (BS1, BS2, BS3, BS4, BS5), one optical switch/ shutter (O-SW), and a 2-D photodetector array.

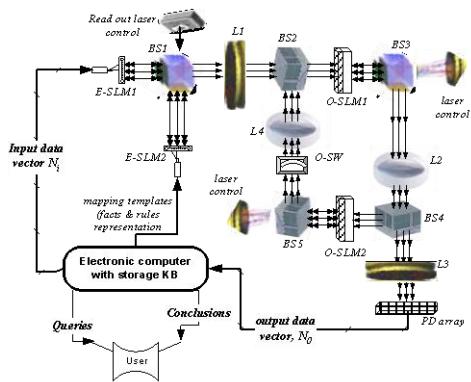


Fig. 8. The complete prototype of the mapping templates-based optical parallel corpus and text mining machine

V. EXPERIMENTAL RESULTS

The proposed architecture is capable of performing many different kinds of text search operations. The optical part constantly performs comparisons while the electronic part is responsible for writing the necessary search arguments, manipulating and interpreting the signals generated by the optical text processor, and passing the results to the user. The following is a description of the implementation of various text operations.

A. Count the Number of Occurrences of a Word in a Document

As an example, the search argument **■Egypt■**, is loaded on SLM-2. The Query Resolver is notified about the arrival of a new document at the OTP and resets a counter (N) to 0. Every new occurrence of "Egypt" will cause an increment of the counter (N) until the end of the document is reached.

This will be used for the document parallelism matching in the parallel corpus as follow:

Content-based matching between any two parallel documents must contain some token pairs that are exact translations of each other. These token pairs are known as translational token pairs. For example, in the two sentences:

"Egypt marks the 38th anniversary of the October victory"
and **،نصر تشهد الذكرى الـ 38 لنصر أكتوبر ،**

the translational token pairs are: ("Egypt" - "مصر"), ("anniversary" - "ذكرى"), ("October" - "أكتوبر"), and ("victory" - "نصر"). Therefore, a simple method to calculate

parallelism for a pair of documents is to search them for translated token pairs, calculate their number as mentioned above, and then use the number of translated pairs found as a value of parallelism similarity. However, a pair in which the position of the two translational token pairs is far from each other is rarely to be a correct translation pair. In the approach proposed here, a reliable threshold θ_d is set to conclude whether a pair of documents is a translation pair or not. Thus, for each English document, the Arabic documents are scanned until one pair with similarity exceeding θ_d is found. The measure of similarity between a pair of documents (A, E) represented by θ_d can be determined empirically as:

$$Sim(A, E) = \frac{2N}{\sum_{A,E} \text{number of tokens}} \quad (1)$$

where N is the number of translational token pairs found between A and E. It is assumed that the difference in position between a good pair of paragraphs varies from (1) to (-1). Therefore, the number of translation pairs N between two documents is based on the total number of translation pairs between paragraphs n_k . Each English paragraph $p_{e,k}$ will be compared to its 3 Arabic paragraphs neighbors $p_{a,k-1}$, $p_{a,k}$, $p_{a,k+1}$, and the one with the maximum value of translation pairs n_k together with $p_{e,k}$ will form a translation pair of paragraphs.

$$N = \sum n_k \quad (2)$$

The technique should use an English-Arabic dictionary of stemmed words. Translational token pairs are found by first stemming the English words with Porter Stemming Algorithm [26], and then looked up in the dictionary for all possible Arabic words.

B. Search for a Word and its Derivatives in a Document, Paragraph or Sentence

The search argument will be **■A■**. For example **■Terrorism■**. As soon as a match is detected the current component will be marked and kept in the buffer for further processing. The arrival of a new component will be detected by keeping track of matches. For the Arabic part of the proposed machine the preprocessor module shown in Fig. 3 is used to divide the Arabic documents into paragraphs, each paragraph is analyzed to extract its keywords by applying the search for word algorithm, and then it extracts the root of each keyword using a morphological module. The output of this module is semi-structured documents in the form of an indexed database. An example is shown in Fig. 9. The first level uses the extracted roots to point to the extracted keywords of a certain root. In the second level index, keywords in turn point to the extracted paragraphs.

The morphological module extracts the root of the each keyword in the issued query on the SLM-1. OSLM-2 then generates all the possible words that have the same root. The optical text miner then generates a group of these derivatives, which have the same semantic features as the original keyword; then the query resolver adds the selected keywords to the keywords list. This will increase the number of resultant paragraphs yielding better recall measures. An Arabic dictionary supported with semantic tags is

implemented using an optical semantic network algorithm in [27].

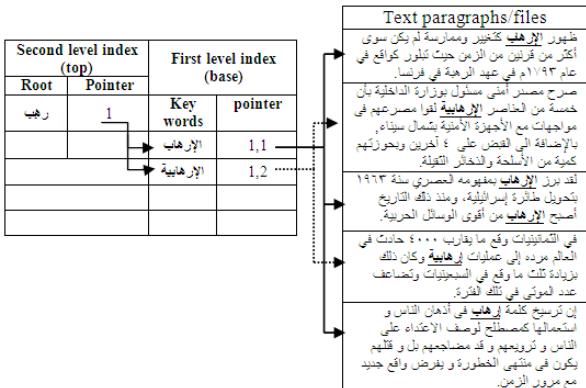


Fig. 9. Semi-structured documents built in a form of indexed database.

C. Search with Query Keywords, Synonyms and Derivatives with the Same Semantic Features

The searching algorithm looks up a Synonym Dictionary for synonyms of the keywords list of the query and adds them to the list. Derivatives of the original list, like the same result of the previous search mode as in Section B, are added to the list as well. The resultant list is used to find the related paragraphs in the database.

D. Search for String A and/or String B in a Document, Paragraph or Sentence

The search argument will be **■A■■■B■** (the number of don't-care characters (**■**) in the middle are arbitrary but have to be at least 1). A match over all the detector cells corresponding to the string A and/or string B will signal a success to the Query Resolver. If the end of the component is reached before the occurrence of the second string (for the "AND" case) the process will start all over again. The following example will help illustrate the procedure. Suppose the user issued a query (Q1) to find all the documents that contain the words "نصر" and "السادات" in the same paragraph and that in one such document (في ذكري نصر أكتوبر يسجل التاريخ للرئيس السادات أنه "صاحب قرار الحرب ويسجل للجيش المصري مجدًا مابعده مجده") in memory of the victory of October) a particular paragraph starts like this:

في ذكري نصر أكتوبر يسجل التاريخ للرئيس السادات أنه "صاحب قرار الحرب ويسجل للجيش المصري مجدًا مابعده مجده"

In English "*In memory of the victory of October history records that President Sadat is the owner of the war resolution and records for Egyptian army later glory*". As previously mentioned, the search argument: **■السادات■■■نصر■** will be loaded in SLM-2 (Fig. 10a) and the appropriate detector cells will be enabled (Fig. 10b). The separation of the two arguments by "don't-cares" is necessary to avoid ambiguities in the detection. The characters of the document begin to slide one-by-one into SLM-1 and are compared to the search arguments.

The query will be satisfied when: 1) a new paragraph is detected, and 2) the presence of strings "نصر" and "السادات" are detected - not necessarily in that order - before the end of the paragraph. When a new paragraph begins, a flag F1 is set to 1 and two others, F2 and F3, are reset to 0 in the Query Resolver. If one of the words is found, the second flag is set. Only when the second word is detected and the three flags register 1, has a match occurred. The paragraph is marked in the buffer and the three flags are reset. The process will be

repeated with the arrival of a new paragraph. Figures 13c-13g show five different instances of SLM-1; namely: the entrance of the paragraph into SLM-1, the detection of the new paragraph, the detection of word "نصر", intermediate shift step and finally, the detection of the word "السادات" which signals the success of the operation.

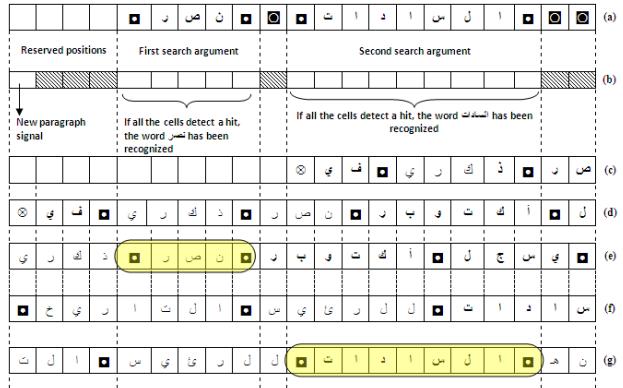


Fig. 10. (a) Contents of SLM-2, (b) the grouping of photodetector cells, (c-g) contents of SLM-1 in five different cycles

The proposed model is tested on an Arabic text book about the history of the Prophet Mohammad (PBUH) named **المرجح المختوم**. Information retrieval systems are usually compared on the basis of the "quality" of the retrieved document sets. This "quality" is quantified using two metrics, Recall ($R = r/K$) and Precision ($P = r/N$), where r is the number relevant and retrieved paragraphs, N is the total number of retrieved paragraphs, and K is the total number of paragraphs in the answer key.

The above algorithm was applied using ten different queries and for each query both the recall and precision were measured when the queries are issued with only keywords, with keywords and derivatives with the same semantic features, and with keywords, synonyms and derivatives with the same semantic features. The overall measured performance in each case and their comparison with fully electronic based results [28] are shown in the Table I.

TABLE I: AVERAGE RECALL AND PRECISION OF TEN QUERIES AND ITS COMPARISON WITH PUBLISHED FULLY ELECTRONIC BASED RESULTS

The other system		The proposed model					
		Keywords only		Keywords+ derivatives		Keywords + derivatives + synonyms	
Recall	Precision	Average recall	Average precision	Average recall	Average precision	Average recall	Average precision
80%	66%	95%	90%	95%	92%	100%	85%

The extreme advantage of the proposed system is the processing time. In our proposed machine the measured average processing time is approximately 32×10^{10} word search operation per second.

VI. CONCLUSION

This paper presented a model and an architectural design for an optoelectronic-based parallel corpus and text mining system based on mapped-template logic and template matching in the optical domain, making full use of the parallelism of optics. The proposed architecture exploits the inherent parallelism features of optics for high speed

database responses and text mining processing. A two-dimensional space optical symbolic correlator was presented based on the VanderLugt correlator to perform the matching operations. Also proposed was an optoelectronic comparison unit that consists of optoelectronic logic, a photodetector array, and the necessary lenses to perform various text operations that can perform approximately 8×10^{10} 32-bit character comparisons per second.

The proposed parallel corpus creation process is executed through mining the web. It is started by classifying English and Arabic web pages, and then creating candidate parallel pairs of documents by filtering them according to their similarity using mapped-template logic and template matching in the optical domain. Finally, the technique measured the parallelism similarity between these candidate pairs. The experimental results of the text mining using an Arabic morphological analyzer and synonyms showed a remarkable performance enhancement in both Precision and Recall measures as well as the superiority of the processing time using the proposed optical system over its corresponding electronic systems. The realization of the proposed optical processor using available optical elements was also addressed. The performance evaluation results showed that the proposed systems are valid, reasonable, and promising for parallelization and speed up of search, indexing and matching.

REFERENCES

- [1] M. Salib *et al.*, "Silicon photonics," *Intel Technol. J.*, vol. 8, pp. 144-160, 2004.
- [2] R. Soref, "Silicon photonics technology: past, present and future," in *Proc. SPIE*, vol. 5730, pp. 19–28, 2005.
- [3] A. Kodi and A. Louri, "Reconfigurable and adaptive photonic networks for high-performance computing systems," *Appl. Opt.*, vol. 48, no. 22, pp. E13–E23, 2009.
- [4] H. J. Caulfield, S. Dolev, and W. M. J. Green, "Optical high-performance computing: introduction to the josa and applied optics feature," *Appl. Opt.*, vol. 48, no. 22, pp. OHPC1–OHPC3, 2009.
- [5] L. Zhao, T. Wang, and S. F. Yelin, "Two-dimensional all-optical spatial light modulation with high speed in coherent media," *Optics Letters*, vol. 34, no. 13, pp. 1930 – 1932, July 2009.
- [6] M. Wang, H. Wang, and Y. Cen, "High-speed digital-image correlation method," *Opt. Lett.*, vol. 34, no. 13, pp. 1955–1957, 2009.
- [7] D. Bimberg *et al.*, "Ultrahigh speed nanophotonics," *SPIE Proc.*, vol. 7211, pp. 721 117.1 – 721 117.10, 2009.
- [8] D. Geer, "Silicon optics aims to combine the best of both worlds," *IEEE Computer Society*, pp. 16 – 19, 2006.
- [9] G. Eichmann, H. J. Caulfield, and I. Kadar, "Optical artificial intelligence and symbolic computing: an introduction," *Appl. Opt.*, vol. 26, no. 10, pp. 1827–1827, 1987.
- [10] R. Gil-Otero, C. J. Moir, T. Lim, G. A. Russell, and J. F. Snowdon, "Free-space optical interconnected topologies for parallel computer application and experimental implementation using rapid prototyping techniques," *Opt. Eng.*, vol. 45, no. 8, pp. 402.1 – 402.6, 2006.
- [11] M. E. Thomas. (July 2010). *Volume holographic optical storage nanotechnology*. Colossal Storage Corporation. [Online]. Available: http://colossalstorage.net_home-diskdrive.htm.
- [12] B. Momeni, S. Yegnanarayanan, M. Soltani, A. A. Eftekhar, E. S. Hosseini, and A. Adibi, "Silicon nanophotonic devices for integrated sensing," *J. nanophotonics*, vol. 3, no. 031001, pp. 1 – 23, 2009.
- [13] R. P. Schmid, T. Schneidera, and J. Reifa, "Optical processing on a femtosecond time scale," *Opt. Comm.*, vol. 207, pp. 155-160, 2002.
- [14] J. W. Na, "A parallel optical computer architecture for large database and knowledge based systems," *CIS*, vol. 3314, pp. 934-939, 2004.
- [15] R. Feldman and J. Sanger, "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data," Cambridge University Press, 2006.
- [16] Internet cashing library. *Internet World Users by Language*. (May 31 2011). [Online]. Available: <http://www.internetworldstats.com/stats7.htm>.
- [17] N. Albalooshi, N. Mohamed, and J. Al-Jaroodi, "The Challenges of Arabic Language Use of the internet," 6th International Conference on Internet Technology and Secured Transactions, UAE, 11-14 December 2011.
- [18] M. M. Sakre, M. M. Kouta, and Ali M. Allam, "Automated Construction of Arabic-English Parallel Corpus," *Journal of the Advances in Computer Science*, ACS, vol. 3, 2009.
- [19] Linguistic Data Consortium. *Automatic Content Extraction Program*. [Online]. Available: <http://projects.ldc.upenn.edu/>, last visited August, 2012.
- [20] S. Ghoniemy and O. H. Karam, "An Optoelectronic Expert System for Knowledge Base Systems," in *Proc. the Tenth Intelligent System Design and Applications Conf.*, Cairo, Egypt, Nov. 2010.
- [21] S. Ghoniemy and O. H. Karam, "Performance of an Optoelectronic Expert System for Massively Parallel Knowledge Base Applications," In *Proc. the IEEE International Symposium on Signal Processing and Information Technology*, pp. 108–, 113, December 15-18, 2010, Luxor, Egypt.
- [22] K. M. Geib *et al.*, "Fabrication and performance of two-dimensional matrix addressable arrays of integrated vertical-cavity lasers and resonant cavity photodetectors," *IEEE STQE*, vol. 8, pp. 943-947, 2002.
- [23] P. S. Guilfoyle and W. J. Wiley, "Combinatorial Logic Based Optical Computing," in *Proc. SPIE*, vol. 639-17, April 1986.
- [24] L. Mei, H. H. Can, and J. Yi, "Principle, equipment and experiment of vector-matrix multiplication by liquid crystal array," presented at International Conference on Information Management and Engineering, 2009.
- [25] Holoeye, Holoeye's Spatial Light Modulator systems, online manulas. (2012). [Online]. Available: http://www.holoeye.com/spatial_light_modulators-technology.html.
- [26] M. Porter. (2006). Porter Stemming Algorithm. [Online]. Available: <http://tartarus.org/martin/PorterStemmer>.
- [27] S. Ghoniemy and O. H. Karam, Semantic Network-based Digital Optical Computer Algorithms for Expert Systems and Massively Parallel Knowledge Base Applications.
- [28] F. Harrag and A. Hamdi-Cherif, "UML modeling of text mining in Arabic language application to the prophetic traditions," *The First International Symposium on Computers and Arabic Language*, Riyadh, 2007.



Samy Ghoniemy received his B.Sc. (Excellent with Honor), and M.Sc. degrees in Engineering in 1990 and 1996 respectively. He also attended a three-term condensed program studying optoelectronic courses during 1991 and 1992. He received his PhD degree from the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada and was nominated for the Senate Medal for the outstanding performance award from the same University. Dr. Ghoniemy's primary research interests are optical computing, optical communications, optical/photonics networks and interconnects. He is currently an associate professor and is a Member of the IEEE Computer society, SPIE "The International Society for Optical Engineering" and OSA "Optical Society of America". Dr. Ghoniemy published more than 50 research papers in the fields of Computer Engineering and Computer Sciences.



Omar H. Karam obtained his B.Sc. in Electrical Engineering (Communications and Electro-physics) from Alexandria University and his M.Sc. and Ph.D. degrees in Computer Engineering from North Carolina State University. He has extensive professional experience in ICT management and is the author or co-author of more than 50 publications in international scientific journals and conferences. Dr. Karam is currently an Associate Professor and the Program Director of the Faculty of Informatics and Computer Science at the British University in Egypt (BUE). He is on seconndom from the Faculty of Computer and Information Sciences, Ain Shams University, Egypt. Prior to the BUE, Dr. Karam was the Director of the Egyptian Universities Network (EUN). His current research interests are Data Mining and Warehousing, Interconnection Networks, Optical Computing and ICT in Education.