

# A Novel Multilingual Web Information Retrieval Method Using Multiwavelet Transform

**Shawki Abdul Rakib Saif Al-Dubae, Nesar Ahmad**

Department of Computer Engineering,  
Zaker Husain College of Engineering and Technology,  
Aligarh Muslim University, Aligarh-202002,U.P, India.  
shawkialdubae@gmail.com, nesar.ahmad@gmail.com

## ABSTRACT

This paper presents a novel approach based on multiwavelet transform in Web information retrieval. The influence multiwavelet transform on feature extraction and information retrieval ability of calibration model and solve problem of selecting optimum wavelet transform for sentence query entered of any language by internet user was investigated. The empirical results show that the proposed method performs accurate retrieval of the multiwavelet transform than scalar wavelet transform. The aptitude of multiwavelet transform to represent one language (domain) of the multilingual world. Regardless of type, script, word order, direction of writing and difficult font problems of the language. This work is a step towards multilingual (English, Spanish, Danish, Dutch, German, Greek, Portuguese, French, Italian, Russian, Arabic, Chinese (Simplified and Traditional), Japanese and Korean (CJK)) search engine. We consider multiwavelet transform as signal processing tool in Soft Computing (SC) as well.

**Keywords:** Multiwavelet Transform, Multilingual, Search Engine.

## 1 INTRODUCTION

Over last two decades, wavelet domain and its applications are increasing very rapidly. Wavelet transform has become widely applied in the area of pattern recognition, signal and image recognition, compression, denoising [7-10]. This is due to the fact that it has good time-scale (time-frequency or multi-resolution analysis) localization property, having fine frequency and coarse time resolutions at lower frequency resolution and coarse frequency and fine time resolutions at higher frequency resolution.

Therefore, it makes suitable for indexing, compressing, information retrieval, clustering detection, and multi-resolution analysis of time varying and non-stationary signals.

Wavelets (the term wavelet translate from *ondelles* in French to English which means a small wave) [11] are an alternative to solve the shortcomings of Fourier Transform (FT) and Short Time Fourier Transforms (STFT). FT just has frequency resolution and no time resolution which is not suitable to deal with non-stationary and non-periodic signal. In an effort to correct this insufficiency, STFT adapted a single window to represent time and frequency resolutions. However, there are limited precision and the particular window of time is used for all frequencies. Wavelet representation is a lot similar to a musical score that

location of the notes tells when and what the frequencies of tones are occurred [10].

In [1-2], we apply a new direction for wavelet transform on multilingual Web information retrieval. The novel method is converted all the sentence query entered of internet user to signal by using its Unicode standard as shown in Figure 1. The main reason to convert the sentence query entered of internet user to Unicode standard. Unicode is international standard for representing the characters used into plurality of languages. Also, it provides a unique numeric character, regardless of language, platform, and program in the world. Furthermore, it is popular used in internet and any operation system of computers, device to text visual representation and writing system in the whole world.

However, we can say that there is no standard method to select a wavelet function in wavelet transform with signals and types of sentence query entered of languages as well. Some criteria have been proposed to select a wavelet. One of them was that the wavelet and signal should have good similarities. This also appears on our previous results of information retrieval where the type of the sentence query entered of language and wavelet should have good similarities.

In fact, it is a good feature of wavelet transform, but at the same time is problem.

In [3-6], the authors have considered a wavelet

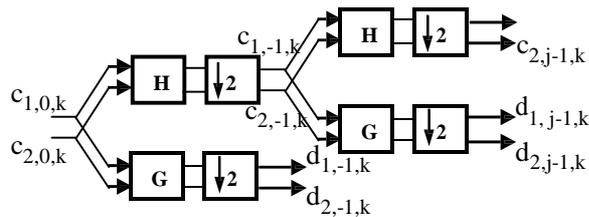


Figure 2: Two levels decomposition in multiwavelet (DMWT).

transform as a new tool in Soft Computing (SC). In [5], the author has proposed adaptive multiresolution analysis and wavelet based search methods within the framework of Markov theory.

In [6], the author has proposed two methods namely, the Markov-based and wavelet-based multiresolution, to search and optimization problems.

The multiwavelet is a body of wavelet transform. Due to, we suggest that consider multiwavelet in SC as well. The main different of wavelet and multiwavelet transforms is that the first transform just has a single scaling and mother wavelet functions and the second has a lot of scaling and mother wavelet functions. The main field of multiwavelet and wavelet transform is used with signals and images. The study and applications of multiwavelet transform is increasing very rapidly which has proven to perform better than scalar wavelet [12-16].

However, we try to investigate our previous suggestion [1-2] to solve problem of selecting optimum wavelet transform for sentence query entered of some languages by internet user.

It is estimated that over 65 percent of the total global online population is non-English speaking. Therefore, the population of non-English speaking internet users is growing much faster than of English speaking users. Asia, Africa, the Middle East and Latin America are the areas with fastest growing online population. Due to, the study of multilingual web information retrieval has become an interesting and challenging research problem in the multilingual world.

Therefore, this is important for using one's own language for searching the desired information. There is a real need to find a new tool to make a multilingual search engine easily.

The 14 languages (English, Spanish, Danish, Dutch, German, Greek, Portuguese, French, Italian, Russian, Arabic, Chinese (Simplified and Traditional), Japanese and Korean (CJK)) belong to the 5 language families. They are selected in this paper because of some reasons. The first reason is that these languages may include the main languages and the most popular language families of the whole world. English, Spanish, Danish, Dutch, German, Greek, Portuguese, French, Italian and Russian languages are Indo-European language families and Arabic, Chinese (Simplified and traditional),

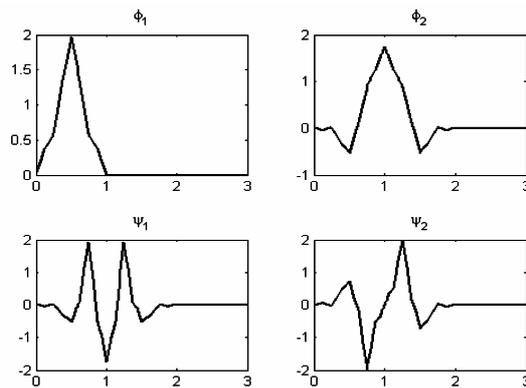


Figure 3: GHM pair of scaling and mother functions of multiwavelets.

Japanese and Korean (CJK) languages belong to Afro-Asiatic, Sino-Tibetan, Japanese and Korean language families respectively. The second reason is that we can easily translate them from English to another language (bilingual) or vice versa by Google™ or other websites.

At the same time, one can obtain much benefit that adopts a multilingual search engine by using multiwavelet transform.

The rest of this paper is organized as follows. Section 2 discusses the preliminaries that are related to multiwavelet transform. In section 3, methodologies pertaining to this work are described. Section 4 contains results and discussions. Finally, Section 5 is for conclusions and future directions.

## 2 An Overview Multiwavelet Transform

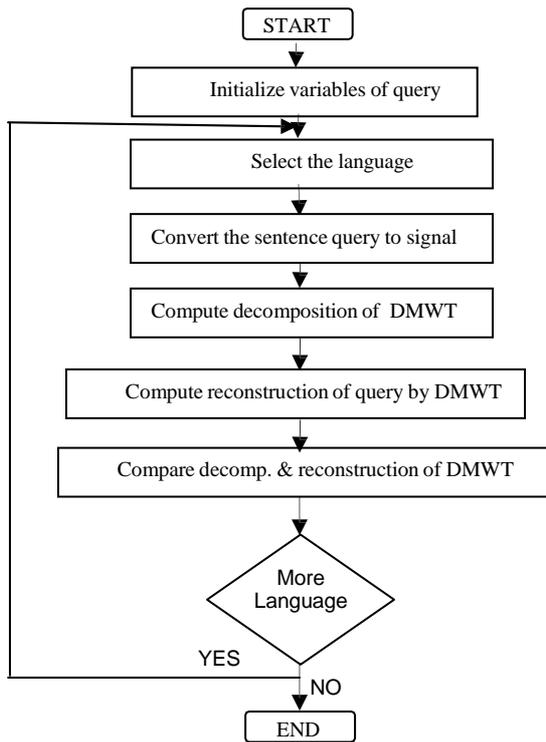
Wavelet transform has a scalar scaling and mother wavelet functions, but multiwavelets have two or more scaling and wavelet functions. The multiwavelet basis is generated by  $r$  scaling functions  $\varphi_1(t), \varphi_2(t) \dots \varphi_{r-1}(t)$  and  $r$  wavelet function  $\psi_1(t), \psi_2(t) \dots \psi_{r-1}(t)$ . Here  $r$  denotes the multiplicity in the vector setting with  $r \geq 2$ . The multiscaling functions  $\varphi(t) = [\varphi_1(t) \varphi_2(t) \dots \varphi_{r-1}(t)]^T$  satisfy the matrix dilation equation [15-16]:

$$\varphi(t) = \sqrt{2} \sum_{k=0}^{M-1} H_k \varphi(2t - k) \tag{1}$$

Similarly for the multiwavelets  $\psi(t) = [\psi_1(t) \psi_2(t) \dots \psi_{r-1}(t)]^T$  the matrix dilation equation is obtain by

$$\psi(t) = \sqrt{2} \sum_{k=0}^{M-1} G_k \psi(2t - k) \tag{2}$$

Where the coefficients  $H_k$  and  $G_k$ ,  $0 \leq k \leq M - 1$ , are  $r$  by  $r$  matrices instead of scalars, which is called matrix low pass and matrix high pass filters



**Figure 4:** Flowchart of proposed information retrieval method. respectively [15-16]. As example, we give the GHM (Geroniom, Hardin, Massopust) system [17-18].

Let parameters filters are as follows:

$$H_0 = \begin{pmatrix} 3\sqrt{5}\sqrt{2} & 4\sqrt{5} \\ -\sqrt{10} & -3\sqrt{10}\sqrt{2} \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 3\sqrt{5}\sqrt{2} & 0 \\ \sqrt{90} & \sqrt{10}\sqrt{2} \end{pmatrix}$$

$$H_2 = \begin{pmatrix} 0 & 0 \\ \sqrt{90} & -3\sqrt{10}\sqrt{2} \end{pmatrix} \quad H_3 = \begin{pmatrix} 0 & 0 \\ \sqrt{10} & 0 \end{pmatrix}$$

and

$$G_0 = \begin{pmatrix} -\sqrt{10} & -3\sqrt{10}\sqrt{2} \\ \sqrt{10}\sqrt{2} & 3\sqrt{10} \end{pmatrix}$$

$$G_1 = \begin{pmatrix} \sqrt{90} & -\sqrt{10}\sqrt{2} \\ -\sqrt{90}\sqrt{2} & 0 \end{pmatrix}$$

$$G_2 = \begin{pmatrix} \sqrt{90} & -3\sqrt{10}\sqrt{2} \\ \sqrt{90}\sqrt{2} & -3\sqrt{10} \end{pmatrix}$$

$$G_3 = \begin{pmatrix} -\sqrt{10} & 0 \\ -\sqrt{10}\sqrt{2} & 0 \end{pmatrix}$$

We have used these parameters with this work.

The original signal  $f(t)$ ,  $f \in L^2(R)$  where  $L^2(R)$  space is the space of all square-integrable functions defined on the real line  $R$  with a basic property, is given by the expansion of multiscaling  $(\varphi_1(t), \varphi_2(t))$  and multiwavelets  $(\psi_1(t), \psi_2(t))$  functions as example depicted in Figure 3 respectively:

$$f(t) = \sum_{k=0}^{M-1} C_j 2^{j/2} \varphi(2^j t - k) + \sum_{k=j_0}^{-1} D_j(k) 2^{j/2} \psi(2^j t - k) \quad (3)$$

$$\text{where } C_j(k) = [c_1(t), c_2(t) \dots c_{r-1}(k)]^T \text{ and}$$

$D_j(k) = [d_1(t), d_2(t) \dots d_{r-1}(k)]^T$  are coefficients of multiscaling and multiwavelet functions respectively and the input signal is discrete. To present the low pass (H) and high pass (G) filters. Based on equations (1) and (2) the forward and inverse Discrete Multiwavelet Transform (DMWT) can be recursively calculated by [15-16]

$$C_{j-1}(k) = \sqrt{2} \sum_k H(k) C_j(2t - k) \quad (4)$$

$$D_{j-1}(k) = \sqrt{2} \sum_k G(k) C_j(2t - k) \quad (5)$$

where the input of data signal should be square matrix. The low pass (H) and high pass (G) filters (previous parameter filters) achieve convolution of the input signal.

Then, each data stream is down sampling by a factor of two as depicted in Figure 2.

### 3 Methodologies

This paper is extended to our previous work [1-2]. The main objective of the proposed method is to make the internet user easily access it using one's language or any language for the required information.

In our previous work, we applied our novel method that the sentence queries entered of 14 languages of five language families by internet user depend on one's own language convert to signal. Then, the signal is processed by using Forty-two wavelet functions (mother wavelets) of six families, namely Haar (haar), Daubechies (db1-10), Biorthogonal (bior1.1- 6.8 and rbior1.1 - 6.8), Cofilets (coif 1-5), Symlets (sym 1 - 10), and Dmey (dmey), of wavelet transform.

However, we can not determine one wavelet function to be used for multilingual web information retrieval perfectly. Due to, the wavelet function, type and sentence query of language should have good similarities. This means that the wavelet function should have good similarities not just with type of language, but also with sentence query of the same language.

In the current paper, we have assumed that some

multilingual (English, Spanish, Danish, Dutch, German, Greek, Portuguese, French, Italian, Russian, Arabic, Chinese (Simplified and Traditional), Japanese, and Korean (CJK)) internet users pose as in the previous our work two sentence queries entered (“What is your name?” and “Good morning, Hi.”). The first sentence query has one capital and small letters and question mark and the second sentence query has small and two capital letters, a comma, and dot.

Therefore, the internet users can pose 16 sentence queries of several different languages (the previous 14 languages with two sentence query with English and Arabic languages so that become 16 sentence queries entered) depend on one’s own language.

The reason of selecting these sentence queries is so as to appraise the accuracy of sentence query reconstruction (retrieval) by DMWT with GHM. Also, we want to further in-depth investigate of some of languages of Indo-European families, the effect of another sentence query with Arabic Language, try to solve problem of selecting optimum wavelet functions of wavelet transform, compare previous and recently results and make on language or transform to multilingual web search engine.

The main reason is to select these languages which may include the features of many language scripts, derivable and the main language families of the world.

As shown in Figure 4, type of language, the first level decomposition is selected, and DMWT is applied to the sentence query entered after convert the query to signal (as matrix) by using Unicode standard. Unicode standard provides a unique numeric character, regardless of language, platform, and program in the world. After that, the approximation and detail coefficients are obtained by using decomposition of DMWT (GHM) of sentence query entered. Then, reconstruction of DMWT is computed. Finally, the average of reconstruction of DMWT has been achieved by comparing the decomposition and reconstruction of every character and space of sentence query entered. These processes are applied to the 16 sentence query entered of all 14 languages of five language families.

Thus, the software is used in order to evaluate retrieval accuracy of one level and using with two dimensions of multiwavelet transform by multiwavelet tools [20].

The normal retrieval result without our developing of multiwavelet tool of the long sentence query entered by internet user “I am a research scholar in Department of Computer Engineering at Aligarh Muslim University, Aligarh, India I am interested in SC” is as follows:

“I am a qerdarcg scholaq hn Departmdns of Compuseq Dnghneeqing `t AkigarhMurkim Uniueqrity, Akigarh, Indh` I am hmserestdd hn RC”.  
Average reconstruction (retrieval) measures as 76%.

Therefore, after our improving the multiwavelet tool under Matlab 7.0 (R14) toolbox [21], the obtained retrieval result is 100%. Due to, we have used our improving multiwavelet tool with evaluating (DMWT) other sentence query of internet user as shown in Table 1 and Table 2.

#### 4 Results and Discussions

The main result is that we can deal with the query entered of language by internet user same as signal within discrete multiwavelet transform or wavelet transform without taking it as image, icon or another method.

In scalar wavelet transform (Fast Wavelet Transform (FWT)), there is no standard method to select a wavelet function with signals and languages. Some criteria have been proposed to select a wavelet. One of them was that the wavelet and signal should have good similarities. We have noticed that the main problem of the sentence reconstruction with scalar wavelet transform, in some of these languages, is reconstructed either different words or shapes other than the words entered or the sentence without space between words and there are different with getting perfect retrieval for Bior2.2 and Bior3.1 of these languages as depicted in Table 1 and Table 2 respectively.

Moreover, the Table 3 and Table 4 show the details analysis of FWT with four levels of Traditional Chinese and Korean languages respectively. Therefore, we have selected level one as good level to get speed up in processing, low size on storage, and quality of retrieval for these languages.

In our previous work, we have suggested to solve problem selecting optimum wavelet function within wavelet transform that use multiwavelet transform.

Due to, the fact that DMWT can simultaneously possess the good properties of orthogonality, symmetry, high approximation order and short support which are not possible in the scalar wavelet transform. The amazing results of discrete multiwavelet transform appear that its retrieval of the information is 100% perfectly with these languages as shown in Table1 and Table2. The accuracy of the reconstructions of DMWT is estimated by average percentage reconstruction of one decomposition level on the 16 sentence query entered of the 14 languages.

This means that Multi-wavelet already has proven to be suitable of multilingual web information retrieval. No matter about type, script, word order, direction of writing and difficult font problems of the language. This means that we can apply all properties of DMWT with GHM parameter filters on text as signal.

However, there is one problem of discrete multiwavelet transform tool which has been improved by us. The problem is that we should enter

the information data as square matrix (two dimensions).

As a result, we need to develop our proposed method with concentrated of decreasing the complexity time of process.

In brief, DMWT has been solved the main problem of selecting optimum wavelet transform for sentence query entered of these languages by internet user. The aptitude of multiwavelet transform to represent one language (domain) of the multilingual world is promising. Regardless of type, script, word order, direction of writing and difficult font problems of the language.

## 5 Conclusions and Future Directions

This paper highlights that applied of multiwavelet transform in Web information retrieval is useful and promising. From the experiments conducted, the evaluation of multiwavelet of solving problem of selecting optimum wavelet function within wavelet transform is to get the multilingual Web information retrieval that is performed using multiwavelet transform with GHM parameters. The results show the satisfactory performance of the applied multiwavelet transform. It is experienced from 16 languages sentences results of 14 languages of five language families and the results show that the multiwavelet with one level give the suitable results for the whole these languages.

In sum, the multiwavelet transform has proved the aptitude for being a multilingual web information retrieval. Regardless of type, script, word order, direction of writing and difficult font problems of these languages. By this property, the power distribution over the multiwavelet domain describes its multilingualism. Also, we improve the multiwavelet tool to obtain perfect results with multilingual web information retrieval.

As a result, we expected that the multiwavelet transform become multilingualism tool on the internet and is a new tool to solve problem of question and answering system in future.

Moreover, we should take care of as further improvement in the proposed work to minimize the computational cost so as to transport it close to the real time process.

## 6 REFERENCES

- [1] S. A. Al-Dubae, and N. Ahmad: New Direction of Wavelet Transform in Multilingual Web Information Retrieval, The 5th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSDK08), IEEE press, October 18-20 , Jinan, Shandong, China, (2008). (to be published)
- [2] S. A. AL-Dubae, and N. Ahmad: The Bior 3.1

- Wavelet Transform in Multilingual Web Information Retrieval, The 2008 International Conference on Data Mining (DMIN'08), a track at The 2008 World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'08), July 14-17 , Las Vegas, Nevada, USA, (2008). (to be published)
- [3] S. Mitra and T. Acharya: Data Mining Multimedia, Soft Computing , and Bioinformatics, J. Wiley & Sons Inc, India, (2003).
- [4] M. Thuillard: Wavelet in Soft Computing. Word Scientific series in Robotics and intelligent systems, Vol. 25, World Sciencfic, Singapore, (2001).
- [5] M. Thuillard: Adaptive Multiresolution and Wavelet Based Search Methods, International Journal of Intelligent Systems, Vol. 19, pp. 303-313, (2004).
- [6] M. Thuillard: Adaptive Multiresolution search: How to beat brute force, Elsevier International Journal of Approximate Reasoning, Vol. 35, pp. 233-238, (2004).
- [7] T. Li, et al.: A Survey on Wavelet Application in Data Mining, SIGKDD Explorations, Vol.4, No.2, pp. 49-68, (2002).
- [8] A. Graps: An Introduction to Wavelets, IEEE Computational Sciences and Engineering, Vol.2, No.2, pp. 50-61, (1995).
- [9] F. Phan, M. Tzamkou, and S. Sideman: Speaker Identification using Neural Network and Wavelet, IEEE Engineering in Medicine and Biology Magazine, PP. 92-101, February (2000).
- [10] R. Narayanaswami, J. Pang: Multiresolution analysis as an approach for tool path planning in NC machining, Elsevier Computer Aided Design, Vol.35, No.2, pp. 167-178, February (2003).
- [11] C. S. Burrus, R. A. Gopinath and H. Guo: Introduction to Wavelets and Wavelet transforms, Prentice Hall Inc, (1998).
- [12] S. Mallat: A theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Trans. On Pattern Analysis and Machine Intell., Vol.11, No.7, pp. 674-693, July (1989).
- [13] G. Strang, and T. Nguyen: Wavelets and Filter Banks, wellesley-Cambridge Press, (1996).
- [14] V. Strela, et al.: The application of multiwavelet filter banks to signal and image processing, IEEE Trans. On Image Processing, Vol.8, No.4, pp. 548-563, April (1999).
- [15] V. Strela: Multiwavelets: Theory and Applications, Ph.D. Thesis, Department of Mathematics at the Massachusetts Institute of Technology, June (1996).
- [16] K. Fritz: Wavelets and Multiwavelets, Chapman & Hall/CRC, (2003).
- [17] G. Plonka, V. Strela: Construction of multi-scaling functions with approximation and

symmetry, SIAM J. Math. Anal., Vol. 29, pp. 482–510, (1998).

- [18] J. Pan, L. Jiaoa, and Y.Fanga: Construction of orthogonal multiwavelets with short sequence, Elsevier Signal Processing, Vol.81, pp. 2609-2614, (2001).
- [19] O. Rioul , and M. Vetterli: Wavelet and Signal Processing , IEEE Signal Proc. Magazine ,Vol. 8 ,No. 4, PP. 14-38, Oct. (1991).
- [20] B. Al jewad, Multiwavelet tools, website: <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=11105>
- [21] M. Misiti, et al.: Wavelet Toolbox MATLAB user’s guide Version 3. Mathworks Inc, (2004).

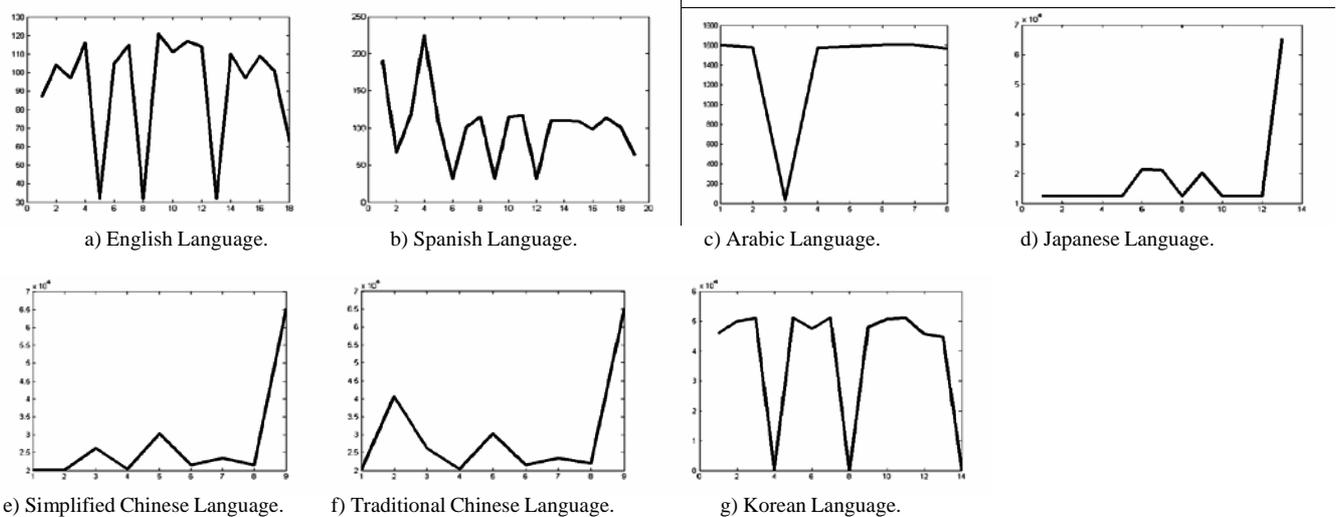


Figure 1: shows “What is your name?” query converted to signal for six languages of five language families.

TABLE 1:

The sentence query of What is your name? translates to 6 Languages with one level decomposition of FWT and DMWT.

The sentence of What is your name? translates to 6 languages(decomposition(D)) and reconstruction (R))							
Compar.	Simplified Chinese	Traditional Chinese	Spanish	English	Arabic	Korean	Japanese
D	什么是你的名字吗?	什麼是你的名字嗎?	¿Cuál es su nombre?	What is your name?	ما اسمك؟	당신의 이름이 무엇입니까?	あなたのお名前は何ですか?
R(DMWT)	什么是你的名字吗?	什麼是你的名字嗎?	¿Cuál es su nombre?	What is your name?	ما اسمك؟	당신의 이름이 무엇입니까?	あなたのお名前は何ですか?
DMWT	100%	100%	100	100%	100%	100%	100%
R(haar)	什么是你的名字吗?	什麼是你的名字嗎?	¿Cuál es su nombre?	Whatisyourname?	ما اسمك؟	당신의이름이무엇입니까>	あなたのお名前は何ですか?
haar	94%	94%	89.5%	83%	87.5%	88%	96%
R (bior2.2)	什么是你的名字吗?	什麼是你的名字嗎?	¿Cuál es su nombre?	What is your name?	ما اسمك؟	당신의 이름이 무엇입니까?	あなたのお名前は何ですか?
bior2.2	100%	100%	100	100%	100%	100%	100%
R (bior3.1)	什么是你的名字吗?	什麼是你的名字嗎?	¿Cuál es su nombre?	What is your name?	ما اسمك؟	당신의이름이무엇입니까?	あなたのお名前は何ですか?
bior3.1	94%	94%	100%	100%	87.5%	84%	96%

TABLE 2:

The sentence query of Good morning, Hi. translates to 10 Languages with one level decomposition of FWT and DMWT.

Sentence of Good morning, Hi. Translate to 10 languages(decomposition(D)) and reconstruction (R))										
Compar.	English	Arabic	Russian	Greek	Dutch	German	Portuguese	French	Italian	Danish
D	Good morning, Hi.	صباح الخير، أهلاً.	Доброе утро, Макс.	Καλημερα, Hi.	Goedemorgen, Hi.	Guten Morgen, Hi.	Bom dia, Hi.	Bonjour, Salut.	Buongiorno, Hi.	Goddag, Hej.
R(haar)	Good morning, Hi.	صباح الخير، أهلاً.	Доброе утро, Макс.	Καλημερα, Hi.	Goedemorgen, Hi.	GutenMorgen, Hi.	Bom dia, Hi.	Bonjour, Salut.	Buongiorno, Hi.	Goddag, Hej.
haar	100%	88.2%	100%	100%	100%	94.1%	100%	100%	100%	100%
R (bior3.1)	Good morning, Hi.	صباح الخير، أهلاً.	Доброе утро, Макс.	Καλημερα, Hi.	Goedemorgen, Hi.	Guten Morgen, Hi.	Bom dia, Hi.	Bonjour, Salut.	Buongiorno, Hi.	Goddag, Hej.
bior3.1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
R (DMWT)	Good morning, Hi.	صباح الخير، أهلاً.	Доброе утро, Макс.	Καλημερα, Hi.	Goedemorgen, Hi.	Guten Morgen, Hi.	Bom dia, Hi.	Bonjour, Salut.	Buongiorno, Hi.	Goddag, Hej.
DMWT	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

TABLE 3:

The sentence query of What is your name? translates to Traditional Chinese Languages with four levels decomposition of FWT (previous work results)

wavelets	Traditional Chinese Language reconstruction question query of 什麼是你的名字嗎？							
	Level 1	%	Level 2	%	Level 3	%	Level 4	%
haar	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
db 1	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
db2	什麻昂你兒同孖嘞 >	61	什麻昂你兒名字嘞 >	72	什麻昂你兒同孖嘞 >	61	億麻昂修兒同孖嘞 >	94
db 3	億麼昂你兒名字嗎？	72	億麼是你兒名字嗎？	78	什麼是你的名字嗎？	89	什麼是你的名字嗎？	89
db 4	什麻是修的同孖嘞 >	67	什麻昂修的同孖嘞 >	61	億麻昂修兒同孖嘞 >	50	億麻昂修兒同孖嘞 >	50
db 5	億麼是你的名字嗎？	83	億麼是你的名字嗎？	83	什麼是你的名字嗎？	89	什麼是你的名字嗎？	89
db 6	億麼昂你的名字嘞？	72	什麼昂你的名字嗎？	89	什麼昂你的名字嗎？	83	什麼昂你的名字嗎？	83
db 7	億麼昂修的名孖嗎？	67	億麼是修的名孖嗎？	72	億麼是你的名字嗎？	83	什麼是你的名字嗎？	89
db 8	億麼是修的名孖嗎？	78	億麼是你的名字嗎？	83	什麼是你的名字嗎？	89	什麼是你的名字嗎？	89
db 9	什麻是你兒同孖嘞 >	67	億麻是你的名字嘞 >	78	什麻是你兒同孖嘞 >	67	億麻昂修兒同孖嘞 >	44
db 10	什麻是你的同孖嘞 >	78	什麻是你的同孖嘞 >	72	億麻昂修兒同孖嘞 >	44	億麻昂修兒同孖嘞 >	44
coif 1	億麼是修的名孖嗎？	78	億麼昂修兒同孖嗎？	56	億麼是修的名孖嗎？	72	億麼是你的名孖嗎？	78
coif 2	什麻是你的同孖嘞 >	78	什麻昂修兒同孖嘞 >	72	什麻昂修兒同孖嘞 >	56	億麻昂修兒同孖嘞 >	44
coif 3	億麻是你兒同孖嗎 >	56	億麻是你的同孖嗎 >	61	億麻是你的同孖嗎 >	61	億麻是你的同孖嗎 >	61
coif 4	什麻昂修的同孖嗎 >	61	什麻是你的同孖嗎 >	72	什麻昂你兒同孖嘞 >	61	億麻昂你兒同孖嗎？	67
coif 5	億麻昂修的同孖嗎？	61	億麻是你的同孖嗎？	66	億麼是你的同孖嗎？	72	億麻昂修的同孖嗎？	67
sym 1	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
sym 2	什麻昂你兒同孖嘞 >	61	什麻昂你兒名字嘞 >	72	什麻昂你兒同孖嘞 >	61	億麻昂修兒同孖嘞 >	44
sym 3	億麼昂你兒名字嗎？	72	億麼是你兒名字嗎？	78	什麼是你的名字嗎？	89	什麼是你的名字嗎？	89
sym 4	億麼是修的名孖嗎？	78	億麼昂修兒同孖嗎？	61	億麼是修的名孖嗎？	78	億麼是你的名孖嗎？	78
sym 5	什麻昂你兒名字嘞 >	67	什麻是你的名字嘞 >	78	什麻昂你兒同孖嘞 >	61	億麻昂你兒同孖嘞 >	56
sym 6	億麼是修的名孖嗎？	78	億麼昂修兒同孖嗎？	61	億麼是修的名孖嗎？	78	什麼是你的名字嗎？	94
sym 7	什麻昂你兒同孖嘞 >	61	什麻是你兒名字嘞 >	72	什麻是你兒同孖嘞 >	67	億麻昂你兒同孖嘞 >	56
sym 8	億麻是修的名孖嘞 >	67	什麻昂修的同孖嘞 >	61	億麻昂修的同孖嘞 >	50	億麻昂修兒同孖嘞 >	44
sym 9	億麼是修的同孖嗎？	72	億麼昂修兒同孖嗎？	61	億麼昂修兒同孖嗎？	61	億麼昂修兒同孖嗎？	61
sym 10	億麼昂修的名孖嗎？	67	億麼是修兒同孖嗎？	61	億麼是你的名字嗎？	89	什麼是你的名字嗎？	94
bior1.1	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
bior1.3	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100
bior1.5	什麼昂修的名孖嗎？	83	什麼昂修的名孖嗎？	83	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
bior2.2	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100
bior2.4	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100
bior2.6	億麼昂修的名字嗎？	89	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100
bior2.8	什麼昂修的名字嗎 >	83	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100	什麼是你的名字嗎？	100
bior3.1	什麼是修的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	89
bior3.3	什麼是你的名字嗎？	94	什麼是你的名字嗎？	89	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94
bior3.5	什麼是你的名字嗎？	100	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100
bior3.7	什麼是你的名字嘞？	89	什麼是你的名字嗎？	94	什麼是你的名字嗎？	94	什麼是你的名字嗎？	100
bior3.9	什麼是你的名字嘞？	89	什麼是你的同孖嘞？	83	什麼是你的同孖嗎？	89	什麼是你的同孖嗎？	78
bior4.4	什麻是你的同孖嘞 >	78	什麻昂修兒同孖嘞 >	61	什麻昂修兒同孖嘞 >	50	億麻昂修兒同孖嘞 >	44
bior5.5	億麼昂修兒名字嗎？	67	億麼是你兒名字嗎？	78	什麼是你的名字嗎？	89	什麼是你的名字嗎？	89
bior6.8	什麻是你的同孖嘞 >	78	什麻昂修兒名字嘞 >	78	什麻昂修兒同孖嘞 >	50	億麻昂修兒同孖嘞 >	44
Dmey	什麼是你的同孖嘞？	83	什麼是你的同孖嘞？	83	什麼是你的同孖嗎？	89	什麼是你的名字嗎？	94
Average %		78		79		79		77



## AUTHORS' BIOGRAPHIES



**Shawki A. Al-Dubaee** was born in 1978, Taiz, Yemen. He received his B.E degree in Computer Technology Engineering from Technical college and M.Sc degree in Computer Engineering from Mosul University, Mosul, Iraq, in 2000 and 2003 respectively with the thesis 'Feature Extraction of Signal Processed Using Wavelet and Artificial Neural Networks'. In 2008, he received his Post Graduate Diploma in Linguistics, Department of Linguistics from Aligarh Muslim University, Aligarh, India. He is a lecturer on leave in Department of Computer and Communication, College of Engineering and Information Technology at Taiz University, Taiz, Yemen.

Currently, he is a research scholar in Department of Computer Engineering, Z. H. College of Engineering and Technology, at Aligarh Muslim University, Aligarh, India. His current research interests include Web Intelligence, Soft Computing, Computational Linguistic, Question and Answering System, and Semantic Web. He is an IEEE student member and Computational Intelligent Society member. Also, he is member in the International Association of Engineers (IAENG).



**Nesar Ahmad** was born in 1961 in Patna, India. He received his B.Sc (Engg) degree in Electronics & Communication Engineering from Bihar College of Engineering, Patna University, India, and M.Sc degree in Information Engineering from City University, London, U.K., in 1984 and 1989 respectively. In 1993, he received his Ph.D. degree from the Indian Institute of Technology (IIT) Delhi, New Delhi, India.

He worked as an Assistant Professor in the Department of Electrical Engineering, IIT Delhi till December 2004. He was with King Saud University, Riyadh during 1997-99 as Assistant Professor. Currently, he is a Professor and Chairman of Computer Engineering Department, Z. H. College of Engineering and Technology, at Aligarh Muslim University, Aligarh, India. His current research interests mainly include Soft Computing, Web Intelligence, and Embedded Systems Design.