Information Streaming Systems: A Review

Jolanta MIZERA-PIETRASZKO,
Institute of Mathematics and Computer Science
Opole University
Opole, Poland
jmizera@uni.opole.pl

Grzegorz KOŁACZEK
Department of Computer Science and Management
Wroclaw University of Science and Technology
Wroclaw, Poland
Grzegorz.Kolaczek@pwr.edu.pl

Ricardo RODRIGUEZ JORGE
Department of Industrial and Manufacturing Engineering
Universidad Autónoma de Ciudad Juárez
Partido Romero, 32310, Mexico
ricardo.jorge@uacj.mx

Edgar Alonso MARTINEZ GARCIA
Institute of Engineering and Technology
Universidad Autónoma de Ciudad Juárez
Partido Romero, 32310, Mexico
edmartin@uacj.mx

***Abstract* — Ubiquitous nature of information determines continuous decision making always based on its stream coming from the environment, not only in the human, but also in the animal world. Communication between the users of information system is a fundamental concept for acquiring, showing, spreading, sharing and constantly increasing the knowledge about the circumstances under which they are in need to make a move all alone as well as while working in groups. In this paper, we review information stream-based systems on amount of information acceptable to facilitate the process of decision making, also in multilingual settings. We analyze the information system objects and the resource management cycle. In addition, methodology of data mining in natural language is investigated. Multilingual information flow between the users is specific while analyzing from the contextual perspective, when the range of information system applications becomes extended to international cooperation.**

***Index Terms* — streaming system, history of Information and Communication Technologies, information system, business intelligence**

I. Introduction

Following Mathematical theory of communication by Claude Shannon [1] – "if a number of messages in a set is finite, then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices are being equally likely". Technological breakthrough signifies the role of information nowadays mainly because the dynamism of transmission and reception determines the quality of the addresser's knowledge. Stream-based systems optimize the process of information flow in real time. Specific instance of such systems refers to multilingual information in which transmission of the input (source language) is different from the reception of the output (target language).

The authors study information processing in streaming systems from the perspective of decision making by the user with the aim at their wide application areas, their classification and their architecture. Our motivation is that as long as there is a lot of literature on information systems as such, streaming systems work mainly in networks, while other application areas seem to be rather neglected.

The remainder of this paper is organized as follows. In section 2, the authors present evolution process of Information and Communication Technologies (ICT) including also communication forms starting from those in the ancient times to forming languages, also international languages, section 3 deals with history of indexing and storing information, including the devices used for that purpose. In the next section we show a great diversity of approaches to defining information systems. Next, the paper moves on to all the parts of information streaming systems, including multilingual streaming systems, then some literature is summarized in brief to highlight the systems' streaming functionalities. Following the topic, we analyze methodology in stream-based processing. The last section summarizes this review.

II. Evolution of Information and Communication Technologies

The essence of information lies in the possibility of making specific decisions, undertakings, planning actions and creating reality on its basis, when considered in general sense, it is a state of social and economic development of the countries.

From historical perspective, communication being one of the forms of information transmission between the humans, dates back to the dawn of time. In the beginnings of humanity, the groups called tribes were created, seeking this way a mutual support for cooperative undertakings, for handing their being, ensuring their security and, above all, the way of existence in the environment of nature.

Historically, the first method of communication was a simple gesture in combination with facial expressions (body language), which over time has been extended by a sound pronounced in a certain tone (connotation). In the next stage of information form development, certain sounds created a word. Systematically increasing number of the newly-formed words turned into simple phrases, followed by expression and then developed to sentences forming a language of the tribe.

In those days people sought for methods of communication in their groups, or more widely in communities, mainly for security reasons, however, it did not mean their willingness to provide information to foreigners. Thus, the process of creating the language of a national group was initiated, which

978-1-5386-5150-6/18/$31.00 ©2018 IEEE

in a broader sense led to the phenomenon of multilingualism referring not only to officially registered languages, but also to slang (informal languages of specific social groups) and dialects (slight grammatical and phonic deformations of the official language) characteristic for the community of a specific territorial region.

Thus, information technology is the science of collecting, memorizing, processing and retrieving information in electronic form. Due to the specificity of multilingual information, the procedures of its processing and searching are different from traditional approaches, because they additionally require taking into account dependencies between the language pairs: source and target.

III. Storage of Information

Information systems become the technological breakthrough of the 1970s, also amongst the individual computer users, as with the emergence of PCs they stimulated enormous development of small and medium enterprises on the business market.

However, the first record of information, found as the preliminary phase of this development process, had been initiated by the ancient Greeks many centuries earlier with the arrival of the papyrus. The original form of indexing information were attached to the rolls of papyrus (see Fig. 1) some strips with titles of texts, thanks to which it was avoided to roll out a whole roll of the papyrus to find an information.

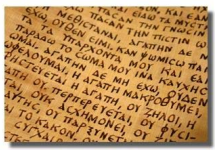

Fig. 1. Ancient Greek manuscript of the 1st Corinthians on a papyrus roll (Source:[1])

Since the papyrus times, almost nothing has been going on until the beginning of the 1950s, when the invention of the computer tape (see Fig. 2) started the era of recording the information in a sequential system, however as yet without possibility of its arrangement or processing. In those days, the technology found the most innovative was random access to the information ensured by a tape which was rolling out from one up to another roll quickly enough to allow the reading device access it right away. Such technology became very popular for faxes when people could transmit their messages to very distant destinations. Just as the first electromechanical computers like ENIAC (Electronic Numerical Integrator and Computer) , Z2 by Konrad Zuse, or Z3 by Turing, or later on an IBM series of computers, the faxes were large enough to occupy the whole room leaving almost no place for the operator.

Fig. 2. Computer tape from IBM Archive Tucson products (Source: [2])

Another way of storing data was a punched tape (see Fig. 3) for teleprinter communication in the early 1960s, also used in telegrams. Such a tape has been widely utilized for a century owning to its rows of holes that allowed to transfer a couple of pieces of information at once to the first computers like Harvard Mark I. Later on, in the 1980s binary data was stored on programmable ROM chips.

Fig. 3. Computer punched tape

The next step in the development of information technology infrastructure was a typewriter (see Fig. 4).

Fig. 4. Ford typewriter from the Martin Howard collection (Source: [3])

According to MIT Press [2] the history of typewriter started in 1873 by Remington Arms company. The creator was Christopher Latham Sholes, the one who redesigned his first sluggish keyboard from alphabetic arrangement to make the most commonly used bars placed at the top left corner. With the beginning of punched tape and typewriters, information was stored on a piece of papers, which initiated a new technology of newspaper and book production.

Next generation of devices for storing information was a hard disc drive launched in 1956 and named IBM 350 disc drive (see Fig. 5) inbuilt in IBM 305 RAMAC computer. Its capacity of information storage was 4.8 MB, which was an equivalent of five million of 7-bit characters. However, it was not the first magnetic device, as its predecessor was a magnetic drum memory 78 RPM records and Edison cylinder.

[1] Reading acts https://readingacts.com/2011/10/31/third-corinthians/

[2] IBM Archive Tucson products http://sysrun.haifa.il.ibm.com/ibm/history/exhibits/tucson/tucson_products.html

[3] Antique typewriters: Martin Howard's collection http://www.antiquetypewriters.com/collection/typewriter.asp?Ford#.WqRpyjeDOUk

978-1-5386-5150-6/18/$31.00 ©2018 IEEE

Fig. 5. IBM 305 RAMAC Computer introduced in 1956 (Source:[4])

In his work of 2004 T. William Olle [3] recommends eight greatest inventions in the history of computing over the period of 50 years:

- Stored program computer hardware and software
- Ferrite core memories hardware
- Higher level languages software
- IBM 360 range hardware and software
- Discs hardware
- Database management software
- Personal computers (PC) hardware and software
- Internet hardware and software

Each of the breakthroughs is just focused on storage of information following the rule, the more information stored on a smaller size of a chip, the better. Thus the list of inventions in computing constitutes simultaneously a comprehensive overview of storing information from the very beginning up to the present time.

IV. Approaches to Defining Information System

Information system seems such a broad term that variety of approaches to defining it lead to the conclusion that it can be seen from different perspectives depending on its application.

In 1999 Nowicki [4] proposed a definition that information system is "spatially distinguished and ordered according to some particular points in time, a set of information pieces, information senders, recipients of information, transmitted by some information channels, whose operation is used to control an economic object such as for instance a company". So someone may deduce that its role is to support organization of company as the information is collected, stored, processed for the purpose of business processes such as management of delivery, marketing and for monitoring supply and demand.
Somehow more multidimentional viewpoint was suggested by Kisielnicki et al [5] who sees it as a matrix of system attributes

$$IS=\{P, I, A, T, O, M\} \quad (1)$$

where;
IS – information system
P – a set of entities that are the system users
I – a set of information about the states and changes occurring in them (system resources)
A – a set of processing algorithms (procedural resources)
T – a set of technical tools (hardware and software resources)
O – a set of solutions used in a company management system
M – a set of metadata, or a comprehensive description of the system information resources

Following this definition, any modification introduced by a user is regarded as an iterative transformation of the matrix sets by another point in time t, that is

$$S_t \rightarrow S_{t+1} \equiv \{P_t \rightarrow P_{t+1} \cap I_t \rightarrow I_{t+1} \cap A_t \rightarrow A_{t+1} \cap T_t \rightarrow T_{t+1} \cap O_t \rightarrow O_{t+1} \cap M_t \rightarrow M_{t+1}\} \quad (2)$$

Mutual relations between the sets, which constitute the elements of the matrix, are assigned binary values according to the principle 1 - relation occurs, or 0 - does not occur, or alternatively, some numerical values defining their specificity of the connections.

As another option, information system can be described as "any combination of information technology with human activity aimed at supporting decisions, management and operation" [6]. A specific feature of information systems is the use of information technology for business processes and communication as a form of the information flow between the system components, or the use of system databases to improve the communication process of the users making decisions with regard to a specific management process.

Below cited are some other definitions of information system not only in order to show their disparity, but principally to present their chief attributes found the in research practice the most significant or maybe even dominant.

1) "The IS is what emerges from the usage that is made of the IT delivery system by users (whose strengths are that they are human beings, not machines). This usage will be made up of two parts:
   (a) First the formal processes, which are currently usually assumed to be pre-determinable with respect to decisions about what IT to use.
   (b) Second, the informal processes, which are what the human beings who use the IT and the formal processes create or invent in order to ensure that useful work is done." [7]
2) "'The information system or management information system of an organization consists of the information technology infrastructure, application systems, and personnel that employ information technology to deliver information and communication services for transaction processing or transaction operations and administration/management of an organization. The system utilizes computer and communications hardware and software, manual procedures, and internal and external repositories of data. The systems apply a combination of automation coming human actions and user machine interaction.'[8]

[4] CED In History of Media Technology by CED Magic
http://www.cedmagic.com/history/ibm-305-ramac.html

3) "An organizational system that consists of technical, organizational and semiotic elements which are all re-organized and expanded during ISD [information system development] to serve an organizational purpose.' [9]
4) "A group of components that interact to produce information. [The five components of an IS are hardware, software, data, procedures, and people.] [10]
5) "Interrelated components working together to collect, process, store, and disseminate information to support decision making, coordination, control, analysis, and visualization in an organization." [11]
6) "Assumed to mean computer-based systems, which are combinations of hardware, software, and telecommunication networks that people build and use to collect, create, and distribute useful information.' [12]

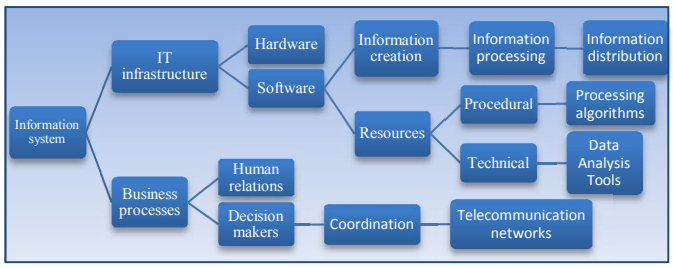

Fig. 6. Information system architecture based on the definitions cited

Fig. 6 shows an IS architecture. To sum up the definitions, the fundamental SI attributes indicated by the authors include: IT infrastructure, business processes, human relations aimed at decision making while creating or processing information, metadata, solutions and the resources such as procedural, or technical tools.

V. Stream-based Information Systems

Section V highlights the streaming information systems' types and their application areas. First and foremost, streaming refers broadly to any information regardless of its form such that we review the state-of-the-art of streaming media, text streaming, video streaming, audio streaming, acoustic streaming, data streaming and text streaming.

Continues signal transmitted to the receiver has numerous advantages predominantly:

- It allows data collection, recording or its usage
- No need to download data
- It saves storage space of the user's device
- High velocity of signal transmission
- Flexibility in delivering information
- Update of the information portions on demand
- Wide range of applications

However, streaming systems have also some limitations that constitute challenges for the research community such as:

- Integration with operational system to make the streaming processor sufficient for retaining big data such as the flow of queries
- Management of stream-based system memory [13]
- Storage costs of the multiple data copies as otherwise it impedes the transmission rate
- Limitation of no more than 1,000 channels per Office group for the subscription price by Microsoft Stream[5]
- Queuing delays in transmission which impact the quality of some stream-based services like Youtube

A. Streaming Media

Streaming Media allows the user to stream any broadcast or media news in real time (live streaming) or on-demand from the preliminarily defined stations depending on the device or the operational system. Some of the most popular standard multimedia coding schemes are MPG-4, 3GP, 3G2, GZ, LFV, FLAG, DiVX, AVI, BIN, 7z, ASX, ASF, AMR, AAC. Several works include mobile learning to fit the student individual learning perception on demand. The authors prove that video streaming media based on cloud mobile learning increase the students' interest in acquiring the knowledge irrespective of their learning style that is whether audio, visual or kinesthetic [14]. Pang Xiyu et al. [15] propose an algorithm for P2P streaming media segmentation aimed at improving the video coding. Quite specific application of streaming media is proposed by Chih-Lin Hu et al. [16]. Their work is focused on an in-car distribution media content to improve visibility of the car driver while on the road. Streaming can be used for alerting when disasters occur, such that A. Kumar Layek et al. [17] explore image stream from social media to identify emergency events. Smart cities can be supported by stream-based media received from collaborative terminals [18]. These are only some of very many examples of the streaming media applications.

B. Acoustic Streaming

Audio Streaming has as long as interesting history. In 1994 during IEEE Ultrasonics Symposium in the US, a group of researchers from Duke University [19] presented their clinical and phantom study on how they had used pulses for inducing acoustic streaming in cyst fluid in order to detect cyst size based on the streaming velocity. This innovative techniques distinguishes cystic from the solid lesions. A model of High Pressure Sodium Lamp whose illuminating light is generated by electric power is presented in [20]. Computing acoustic streaming parameters leads to making an approximation of acoustic resonances effect on the lamp. Reynolds streaming is a form of acoustic streaming used for heating dissipation in the cases where ultrasonic active cooling devices are not effective. J. Löschke et al. applied Reynolds streaming [21] for thermal management. Their cooling system got the loss from the driver electronics of less than 10%. Weighted audio visual information streams are estimated based on entropy to their reliability for speech recognition [22].

$$H_s(t) = -\sum_i P_{is}(t) log_2 P_{is}(t) \quad (3)$$

[5] Microsoft Stream: Quotas & Limitations https://docs.microsoft.com/en-us/stream/quotas-and-limitations

Their entropy $H_s(t)$ for stream s in time t are estimates of the confidence, where $P_{is}(t)$ is a prior probability in a previous state according to Bayes rule over all the states being vocabulary words. Their findings proved promising on the comparison basis to other methods in literature. Early detection from the streaming audio is proposed in [23] as opposed to other approaches aimed at detection of the complete events. The method uses random forest framework. Their system identifies all categories of the events recorded. Totally different application area is suggested by using a new generation technology Visible Light Communications (VLC) for processing acoustic data streams [24, 25]. Compared were PCM and MP3 streaming over 20 cm error free VLC channel In the both cases, the audio stream quality increases proportionally to the higher data rates [24].

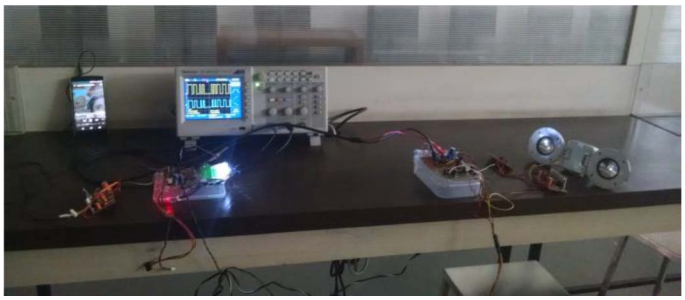

Fig. 7. Complete hardware setup, Source [25]

Fig. 7 shows the audio stream processing starting from the mobile phone audio input being amplified and digitized, then the audio data portions are sent through the controller to the VLC module (the shinning light in Fig. 7). After that, the audio data is extracted in the receiver module and transferred to the speaker [25].

C. Video Streaming

The greatest challenge of video streaming in general is bandwidth consumption, this applies in particular to mobile services. Ch. G. Bampis et al. [26] use recurrent neural networks for real time prediction of quality of experience. Adaptive video streaming enables addressing the users' needs with heterogeneous expectations. S. Issa et al. [27] adopt Enhancement Fractional Participative Scheme implementing JSVM 9 14 environment for evaluation of the low bandwidth. Transmission of video quality is also studied in vehicle driving when switching between the RSUs. Cloud-based predictive model explores quality of the video portions in the RSUs and predicts it based on the data gathered [28]. A bit similar approach to quality analysis based on linear and support vector regression machine learning algorithm but referring to H264 streams over wireless networks is proposed in [29]. The results obtained were very promising as the accuracy reached more than 90%. Y. Tian et al. [30] show how to improve Amazon S3 pricing in cloud-based video streaming on a mobile device. The video encoded segments go to the Amazon S3 bucket to enable the user to retrieve the most recent video and thus to avoid payment for the idle time of the server.

D. Streaming Data

As data refer to variety forms of information and to a huge amount of content, they are strongly associated with research studies and explored from different perspectives. For instance, partial discharges in transformers can be precisely localized based on massive high-frequency streaming sensor data. A voltage threshold has been set to obtain 400ns signal samples from the data stream which is being timed for identification of partial discharges in transformers [31]. Streaming data can be modeled to ensure equitable Poisson distribution for tracking and recovering information from real video data. Wang et al. [32] propose a following stochastic approach: based on some observations $\{y\}_{n-1}^{M}$ they estimate subspace matrix $\{\widehat{D}\}_{t-1}^{M}$ of a video data stream to be tracked and recovered, then they compute a coefficient vector $\hat{a}_t$

$$\hat{a}_t = \underset{a \in R_+^k}{\operatorname{argmin}} -logPois(y_t : \{\widehat{D}\}_{t-1}^{M} a) + \mu\|a\|_2^2 \quad (4)$$

Following, each row of the subspace matrix $\{\widehat{D}\}_{t-1}^{M}$ is iteratively updated by the projected gradient descent

$$\hat{d}_{t,i} = \underset{a \in R_+^k}{\operatorname{argmin}} -\frac{1}{t}\sum_{n-1}^{t} logPois\,(y_{n,i} : \{\widehat{d}\}_i^T \widehat{a_n}) + \mu\|a\|_2^2 \quad (5)$$

D. Marciani et al. [33] apply data streaming to support law enforcement agencies with investigation of criminal networks. Making assumption about the criminals' interest in using the same links, they design a pattern of hidden and potential links with the aim at extracting the information essential for investigation of criminal cases.

E. Text Streaming

Social media provide enormous amount of text streaming. Comparison of effectiveness in identifying sarcasm from the short tweets of no more than 160 characters is made based on the following algorithms: Random Forest, Gradient, Boosting, Decision Tree, Adaptive Boost, Logistic Regression and Gaussian Naïve Bayes [34]. The tweets were classified manually, as a result the most efficient algorithm turned out to be AdaBoost, the second best Gradient Boost, then Random Forest. Another approach to tweets is discussed making use of velocity of those appearing in social media and the search terms to access historic tweets. The authors' findings indicate that the relevance increases proportionally to the best query terms [35]. Clustering events based on short text streams is studied. In order to measure a distance between the events the authors of [36] build microclusters from the stream summery. For the parameters such as time location or binary signature applied is the following similarity measure

$$Sim(t_1, t_2) = \frac{1}{N}\sum_i w_i \times score(i) \quad (6)$$

where $Sim(t_1, t_2)$ is similarity between two topics while w_i are their weights. N is the number of elements in *t.* Quite well-

known algorithm B25 on the other hand, is used for ranking the contents of the public opinion platform. After filtering spam portions of the text, the topics are clustered [37].

F. Image Streaming

While the truth is that processing images requires artistic skills, the real artifact involves integration of the graphical attributes such as color, chroma, texture, photo retouching, shapes, resizing objects, canvas, shadows, gradients, contours, transparency, or lines. These are found the parameters of image streams. Analytical geometry is typically applied to solve linear transformation of graphical elements such as point (pixels), lines (edges) or to reduce dimensionality of spaces (plane) aimed at improving image quality. Image content analysis relies on image annotation semantics, such that Google Cloud Vision API helps limit the time of getting the annotation while the WordNet supports mapping of the undefined terms [38]. An agriculture literature review presents classification of fruit images processed with the usage of the techniques such as enhancement of the features selected like contrast or brightness, colour conversion model (HSV, HIS, HSL), image segmentation for an object detection and image content analysis [39]. Researchers from India study medicinal image qualities extracted from a database including 75 herbal leaves. Their algorithm processes contrast, correlations of pixels, uniformity of energy, homogeneity. The image texture is studied based on the qualities like "perceived lightness, uniformity, density, roughness, regularity, linearity, frequency, phase, directionality, coarseness, randomness, fineness, smoothness, granulation" [40]. Face recognition is a robust when applied is Scaled Invariant Feature Transform and Multidirectional Multidual Cross-Patterns techniques. The parameters include illumination, pose, occlusion and expression [41]. Remote Sensing Image Processing technique uses data from numerous satellites. Compared are multispectral and panchromatic images [42] Image gray scale and smoothing are used for enlightening arrester' s action in converter station. Neural network is trained for learning the arrester position patterns [43]. Monitoring highway traffic for limitation of the motorway blockades is analyzed using image processing technique that consists of the following steps: object recognition, object tracking, zone and anomaly detection, morphological operation aimed at identification of a vehicle position, direction of the vehicle motion and its lane. According to the authors the methodology proposed provides very promising results regardless of the day or night time [44].

VI. MULTILINGUAL INFORMATION STREAM-BASED SYSTEMS

While browsing information retrieval engines, the user's need to access information in languages different from the query language is determined by some factors, such as:

- Analysis of the state of research and the projects currently implemented all over the world in a specific scientific domain for instance in order to establish international cooperation
- Passive knowledge of the resulting documents' language, however, sufficient enough to read technical documentation of various types
- The database contains the documents in such a large number of languages that it is impossible to formulate a query in any language, while learning about their content is facilitated by a built-in computer translation module
- A need to contact a specialist from another country for advice, or for medical help
- A database of the images which are annotated in a foreign language footnotes that are unclear for the user
- The user is not equipped with enough language resources to help him understand the information

Accuracy of the preliminary diagnosis is determined during the interview between GP and patient. It seems a challenge for a native speaker to understand medical terminology, let alone for a foreigner. An approach to comprehensive evaluation procedure of e-Health language-based services offered to multinational patients [45] involves lexical density *LD* and Fox Index *FI*

$$LD = \sum_{i} \frac{w_i}{N_w} \times 100[\%] \tag{7}$$

where w_i denotes medical term being a hard word token and N_w represents the number of all word tokens, whereas

$$FI = (\sum_{s} w_s + w_{3s}[\%]) \times 0.4 \tag{8}$$

where w_s denotes an average number of word tokens in a sentence and w_{3s} a percent number of at least 3-syllable word tokens. Out of 7 evaluated EU multinational medical projects, only two stepped across the low threshold of *LD*=10,6 (comprehensibility of a junior high school) and *FI*=30%.

Computer-Assisted Multilingual Decision Support system [46] defines information flow in real time between medical professional and patient by providing suggestions of the possible set of diseases anticipated based on the symptoms the patient is admitted to present. The number of elements in the diagnosis set decreases in time proportionally to the information flow resulting in the final diagnosis. It has been found by medical professional the most essential at the emergency hospital units.

Mapping source to target languages [47] is an innovative MT technology published by the Polish Patent Office. The system produces an equivalent translation of the input data stream. Source-Target Mapping (STM) model is grounded upon reversibility of the translation algorithm which incorporates symmetric relation and transitive law. In addition, the paper presents a couple of case studies to provide an evidence of how it works in practice.

Social media automatically translated filtering of Japanese tweets is comparable to human translation [48], even more, it is faster and cheaper with online Google Translator and Bing

Translator. Nonetheless, the authors missed the neural network technology introduced by Google in 2016 that produces human-like translation results since the both systems are trained in real time on all the information entered all over the world to any services offered by Google making it the almost perfect tool. "*Almost*" here refers to the principle – the more text entered in a particular source language, the closer to human translation result can be achieved. Another principle – the shorter text such as social media tweets, the better translation as it includes less grammar and usually a lack of anaphora or cataphora.

Multilingual Phone Recognition of phonetic units language independent system is proposed in [49]. The system was tested on 4 Indian languages for recognition of automatic language detection of the read radio or TV broadcast news or books by a speaker. The authors applied Hidden Markov Model, Fourier Transform and Viterbi decoder.

Optimization of a query entered to multilingual search engine is discussed by the authors of [50]. Their system parses the query to select a plan which produces the minimal cost of relations, projections, joints or indexes. Compared were English and Gujarati languages based on 50 records. The minimal cost of the Gujarati queries was 51 while the English queries produced the cost of 25. According to the authors, such a result indicates the importance of query optimization in their Gujarati language.

UNL Interlingua (Universal Networking Language) is a question answering system for processing factoid questions with question word such as *why, which, what* and *where* [51]. An input is automatically transformed by interactive analyzer into Interlingua representation. The EUGENE tool supports the output with linguistic dictionary to make it natural language. Other components complete the output with the missing words. The accuracy in answering 37 questions was 82.2%.

VII. Methodology in Stream-Based Processing

This section summarizes the methodologies adopted by the research in streaming systems depending upon the data form.

TABLE I. Methodologies in Processing Stream-Based Information

Stream type	Streaming systems		
	Methodology	***Application***	***Reference***
Media	Machine learning	Cloud mobile learning	[14]
	Media segmentation	Video coding	[15]
	Distribution of media contents	Car drive	[16]
	Collaborative terminals	Smart cities	[18]
Audio	Signal clustering	Cyst size	[19]
	Acoustic resonance	HPSL	[20]
	Weighted values on entropy	Speech recognition	[22]
	Random forest	Early detection	[23]
	Amplification & digitization	VCL	[24, 25]
Video	Recurrent neural networks	Quality of experience	[26]
	Fractional adaptive method	bandwidth	[27]
	Switching between RSUs	Prediction	[28]
	Linear support vector regression	Wireless networks	[29]
	Video segment encoding	Amazon S3 pricing	[30]
Data	Sensor signal samples	Discharges	[31]
	Poisson distribution	To recover inform.	[32]
	Pattern of hidden links	Criminal networks	[33]
Text	Gaussian Naïve Bayes	Sarcasm	[34]
	Velocity of terms	Best query terms	[35]
	Similarity	Clustering events	[36]
	B25	Ranking opinion	[37]
Image	Google Cloud Vision API	Annotation semantics	[38]
	Image segmentation	Object detection	[39]
	Contrast, correlations of pixels, energy, homogeneity	Medicinal qualities	[40]
	Invariant Feature Transform	Face recognition	[41]
	Remote Sensing from satellites	multispectr panchroma	[42]
	Image gray scale & smoothing	Arrester' s action	[43]
	Motorway blockades	Vehicle position	[44]
Multi-lingual	Lexical density and Fox Index	Medical diagnosis	[45]
	Modeling symptoms	Decision Support	[46]
	Source-Target Mapping	Machine Translation	[47]
	Google vs Bing Translator	Inform. Filtering	[48]
	Hidden Markov Model	Language detection	[49]
	Query parsing	Optimal querying	[50]
	Universal Networking Lang.	Question answering	[51]

Table I presents state-of-the-art in processing stream-based systems and their applications. It may lead to a conclusion that the widest range of practical solutions is given by machine learning methods and similarity measures. Some other popular approaches in research include Hidden Markov Model and mapping texts for processing Machine Translation. In addition, weighting some attributes and clustering set elements remain preferred for acoustic signal processing.

VIII. Conclusion

This paper provides a comprehensive review of the state-of-the-art based on selection of the newest 51 research works out of a couple of hundreds. We present the approaches to

developing stream-based information systems, their definitions, classification, the backbone attributes and the evolution of information systems, which over around 50 years has created new communication channels for the human society. It shows dynamism of innovative methodologies aimed at extending infrastructure in ITC and its dominant role in our life nowadays.

REFERENCES

[1] C. E. Shannon, "Mathematical Theory of Communication", The Bell System Technical Journal, Vol.27, p.379-423, American Telephone & Telegraph Co, USA, 1948.

[2] R. C. Cassingham, *The Dvorak Keyboard*, (Arcata, California: Freelance Communications, ISBN: 0-935309-10-1), pp. 21-24, MIT Press, 1986.

[3] T. William Olle, "Eight Significant Events in the 50 Year History of Computing", in Proceeding - History of Computing in Education, (Eds) John Impagliazzo, John Lee, IFIP 18th World Computer Congress TC3/TC9 1st Conference on the History of Computing in Education 22–27 August 2004 Toulouse, France, Springer Nature, 2004.

[4] Nowicki A., "Strategy for Improving Information System in Business Management", University of Economics, Wroclaw, Poland, 1999.

[5] J Kisielncki, H. Sroka, "Business Information Systems: Computr Science for Management", Placet, Warsaw, Poland, 1999.

[6] Ellison R., Moore A.: Trustworthy Refinement through Intrusion-Aware Design (TRIAD), Networked Systems Survivability, Technical Report, Carnegie Mellon, Software Engineering Institute, Pittsburgh, 2003

[7] Paul RJ., "Challenges to information systems: time to change", European Journal of Information Systems 16(3), pp.193–195, 2007.

[8] Davis GB.: Information systems conceptual foundations: looking backward and forward. In Organizational and Social Perspectives on Information Technology (Bakserville R, Stage J and Degross JI, Eds), pp 61–82, Kluwer Academic Publishers, Boston, 2000.

[9] Lyytien K. and Newman M.: Punctuated equilibrium, process models and information system development and change: towards a socio-technical process analysis. Sprouts: Working Papers on Information Environments, Systems, and Organizations 6(1), pp.1–48, 2009

[10] Kroenge DM.: Experiencing MIS. Pearson Prentice-Hall, Upper Saddle River, NJ., 2008

[11] Laudon KC and Laudon JP.: Management Information Systems: Managing the Digital Firm 10th Edn. Pearson Prentice-Hall, Upper Saddle River, NJ., 2007

[12] Jessup L and Valacich J.: Information Systems Today: Managing in the Digital World 3rd ed., Pearson Prentice-Hall, Upper Saddle River, NJ., 2008

[13] J. G. Apostolopoulos, W.Tan, S. J. Wee, „.Video Stream: Concepts, Algorithms and Systems", Hewlet Packard, Semantic Scholar, 2002.

[14] C. Riana, Munir, Misrina, "The Influences of Video Streaming Media based on Cloud Mobile Learning in Learning Styles", Proceedings of 3rd International Conference on Science in IT, Universitas Pendidikan, Indonesia, IEEE Indonesia Section, 2017.

[15] P. Xiyu, W. Cheng, X. Zhonghua, "A P2P Streaming Segmentation Method based on Scalable Video Coding Data Characteristics", Proceedings of 2nd International Conference on Mechanical Control & Computer Engineering, Harbor, China, IEEE Computer Society, 2017.

[16] Ch. Hu, P. Chiu, S. Chuang, "CarView: A Live Media Streaming Platform Among Neighbour Vehicles on Road:, Proceedings of International Conference on Applied Innovation Systems, Sapporo, Japan, May, 2012, IEEE Computer Society, 2017.

[17] A.K. Layek, A. Gupta, S. Ghost, S. Mandal, "Fast Near Duplicate Detection from Image Streams on Online Social Media during Disaster Events", Proceedings of 2016 Annual India Conference, Bengaluru, India, IEEE Computer Society, 2017.

[18] J. M. Batalia, P. Krawiec, C. Mavromoustakis, G.Mastorakis, N. Chilamkurti, D. Negru, J. Quyreix, E. Borcoci,"Efficient Media Streaming with Collaborative Terminals for the Smart City Environments" IEEE Communications Magazine, vol. 50, Issue 2, pp. 98-104, 2017.

[19] K. R. Nightingale, P. J. Cornguth, W. F. Walker, A. Jett, G. E. Trahey, "Generation and Detection of Acoustic Streaming to Defferenciate Between Solid and Cystic Beast Lesions", 1994 Ultrasonic Symposium, USA, IEE Computer Society, 1994.

[21] A. Toumi, L. Chhun, S. Bhosle, G Zissis, P. Maussion, J. Hirsh, "Acoustic Resonance Characterization and Numerical Model including Acoustic Streaming in HSP Lamp", IEEE Transactions on Industry Applications, vol. 49, Issue 3, pp. 1154-1160, IEEE Computer Society, 2013.

[22] M. Gurban, J. P. Thiran, "Using Entropy as a Strean Reliability Estimate for Audio Visual Speech Recognition", European Signal Processing Conference, Lausanne, Switzerland, IEEE Computer Society, 2008.

[23] H. Pha, E. Maass, R. Mazur, A. Mertins, "Early Detection in Audio Stream", IEEE International Conference on Mulimedia and EXPO, Torino, Italy, IEEE Computer Society, 2015.

[24] D. Samudika, L. Jayasinghe, K. Gunathilaka, Y. Rumesh, R. Weerasuriya, D. Dias, "Stereo Audio Streaming via Visible Light", IEEE Moratuwa Engineering Research Conference, Sri Lanka, IEEE Computer Society, 2016.

[25] M. S. Gujar, S. Velankar, A. Chavan, "Real Audio Streaming using Visible Light Communication", International Conference on Inventive Computation Technologies, Coimbatore, India 2016, IEEE Computer Society, 2016.

[26] Ch. G. Bampis, Z. Li, I. Katsavounidis, A. C Bovik,"Recurrent and Dynamic Models for Predicting Streaming Video from Experience", IEEE Transactions on Image Processing, vol. PP. Issue 99, IEEE Computer Society, 2018.

[27] S. Issa, O. O. Khalifa, T.S. Gunawan, "Analysis for Adaptive Video Streaming for Users with Demand Heterogenity", Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), November 2017, Putrajaya, Malaysia, 2017.

[28] E.B. Smida, S.G. Fantar, H. Joussef, „Predictive Handoff Mechanism for Video Streaming In a Cloud-Based Urban Vanet", 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications, Tunisia, IEEE Computer Society, 2017.

[29] Ch. Singhal, G.T. Tadepalli, „Machine Learning Based Subjective Qualty for Video Streaming over Wireless Networks",20th International Symposium on Wireless Personal Multimedia Communications (WPMC2017), Indonesia, IEEE Computer Society, 2017.

[30] Y. Tian, R. Babcock, C. Taylor, Y. Ji,"A New Live Video Streaming Approach Based on Amazon S3 Pricing Model", 8th Annual Computing and Communication Workshop and Conference, Las Vegas, USA, IEEE Computer Society, 2018.

[31] J. Wang., K Wu, A. Sim, S. Hwangbo,"Accurate Signal Timing from High Frequency Streaming Data", 2017 IEEE International Conference on Big Data (BIGDATA), 2017 Boston, USA, IEEE Computer Society, 2017.

[32] L. Wang, Y. Chi,"Stochastic Approximation from Memory-Limite Subspace Tracking for Poisson Streaming Data", IEEE Transactions on Signal Processing, vol. 66, No. 4, FEB, 15, 2018. IEEE Computer Society, 2018.

[33] G. Marciani, M. Porretta, M. Nardelli, G.F. Italiano,"Data Streaming Approach to Link Mining In Criminal Networks", 2017 5th International Conference on Future Internet of Things and Cloud Workshops, Prague, Czech Republic, IEEE Computer Society, 2017.

[34] A.G. Prasad, S. Sanjana, S. Bhat, B.S. Harish,"Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data", 2017 2nd International Conference on Knowledge Engineering and Applications, London, UK, IEEE Computer Society, 2017.

[35] N. Bhattacharya, I. B. Arpinar, U. Kursuncu,"Real Time Evaluation of Quality of Search Terms during Query Exansion for Streaming Text Data Rusing Velocity and Relevance", 2017 IEEE 11th International Conference on Semantic Computing, San Diego, USA, IEEE Computer Society, 2017.

[36] Z. Sun, J. Han, H.W. Hao,"Strucural Feature-Based Event Clustering forShort Text Stream", 23rd International Conference on Pattern Recognition (ICPR), Mexico, 2016, IEEE Computer Society, 2016.

[37] C. Yuan, H. Lin, X. Shang, Ch. Liu, L. Wang,"OPO: Online Public Opinion Analysis System over Text Streams", International Conference

978-1-5386-5150-6/18/$31.00 ©2018 IEEE

on Service Systems and Service Management, Dallian, China, 2017, IEEE Computer Society, 2017.

[38] S-H. Chen and Y-H Chen, "A Content-Based Image Retrieval Method Based on the Google Cloud Vision API and WordNet", In: Nguyen N., Tojo S., Nguyen L., Trawiński B. (eds) Intelligent Information and Database Systems. ACIIDS 2017. Lecture Notes in Computer Science, vol. 10191. Springer, Cham, 2017.

[39] C.S. Pereira, R. Morais, M. Reis.,"Recent Advances in Image Processing Techniques for Automated Harvesting Purposes: A Review", Intelligent Systems Conference, London, UK, pp. 566-575, IEEE Computer Society, 2017.

[40] T. Vijayashree, A. Gopal, "Leaf Identification for the Extraction of Medicinal Qualities Using Image Processing Algorithm", Proceedings of International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 2017, IEEE Computer Society, 2017.

[41] W. Jebarani, T. Kamalaharidharini, "Robust Face Recognition and Classification System Based on SIFT and DCP Techniques in Image Processing", Proceedings of International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 2017, IEEE Computer Society, 2017.

[42] J. Cai, J.X. Ma, J.X. Wang, "Multispectral Image and Full Color Remote Sensing Image Processing Technology", Proceedings of 9th International Conference on Modeling, Identification and Control (ICMIC 2017), Kunming, China, IEEE Computer Society, 2017.

[43] C. Jiarui, J. Xiaofei, Y. Yue, L. Yejin, J. Shichao, Z. Bo, "A Recognition Method of Action-Times-Readout of Arrester in Convertor Station Based on Image Processing", Proceedings of 9th International Conference on Modeling, Identification and Control (ICMIC 2017), Kunming, China, IEEE Computer Society, 2017.

[44] S. Shashikar, V. Upadhyaya, "Traffic Surveillance and Anomaly Detection using Image Processing", Proceedings of 4th International Conference on Image Information Processing (ICMIC 2017), Waknaghat, India, IEEE Computer Society, 2017.

[45] J. Mizera-Pietraszko, P. Świątek, "Access to e-Health Language Based Services for Multinational Patients", Conference: 17th International Conference on E-health Networking, Application & Services (HealthCom), Boston, MA Date: OCT 13-17, 2015, pp. 232-237, IEEE Computer Society, 2017.

[46] J. Mizera-Pietraszko, "Computer-Assisted Clinical Diagnosis in the Official European Union Languages", Conference: 18th International Conference on E-Health Networking, Application & Services (HealthCom), Munich, MA Date: SEP 14-17, 2016, pp. 241-245, IEEE Computer Society, 2016.

[47] J. Mizera-Pietraszko, G. Kołaczek, R. Rodriguez Jorge, "Source-Target Mapping Model of Streaming Data Flow for Machine Translation", Proceedings of International Conference on Innovations in Intelligent Systems and Applications (INISTA), Gdynia, Poland, pp. 490-495, IEEE Computer Society, 2017.

[48] Q. Hou, A. Musaev, Y. Yang, C. Pu, "A Comparative Study of Increasing Automation in the Integration of Multilingual Social Information", Proceedings of 3rd International Conference on Collaboration and Internet Computing, California, USA, IEEE Computer Society, 2017.

[49] K.E. Manjunath, K.S. Rao, D.B. Jayagopi, "Development of Multilingual Phone Recognition System for Indian Languages", IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Bangalore, India, IEEE Computer Society, 2017.

[50] B. Shah, J. Pareek, "Query Optimization for Information Retrieval in Multilingual Environment for e-Governance Resources", International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, IEEE Computer Society, 2016.

[51] K Goel, P. Bhatia, "Information Retrieval Systems using UNL for Multilingual Question Answering", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India, IEEE Computer Society, 2016.

978-1-5386-5150-6/18/$31.00 ©2018 IEEE