Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of Information Technology and Networking

S.J.I.T.N
Saba Journal of Information Technology and Networking

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of Information Technology and Networking

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of Information Technology and Networking

Contents

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of Information Technology and Networking

Article

OPEN GOVERNMENT INITIATIVES IN PUBLIC SECTOR: A PROPOSED FRAMEWORK FOR FUTURE RESEARCH

Dr.Emad Abu-Shanab
Yarmouk University, Jordan

Article info

Article history:

Accepted Dec, 2014

Keywords:

E-government,

Open government,

Open data

Abstract

The initiatives taken by many governments to open data flows between them and their societies are attracting huge attention by researchers and civil society bodies. The evolution from (OG1.0) to (OG2.0) started by emphasizing transparency and more engagement of the public. To attain such level, governments need to improve online engagement, make engagement meaningful through competitions and challenges, identify new collaborative platforms, build evaluation into program design, and integrate co-production within agencies initiatives. Open government concept did not stop at Obama's administration proposed dimensions, but extended to more complex and complicated models. This paper will examine the concept of open government and its major dimensions, review the literature, visit some experiences of the world, and propose a framework for research.

* Corresponding author: Dr. Emad Abu-Shanab
E-mail address: abushanab.emad@gmail.com

1. Introduction

The United Nations e-government reports indicate that the development and achievements of member nations in relation to the progressive stages proposed is acceptable in the first and second stage. The 2012 report merged the first two stages, where the existence and information stage were merged into one stage labeled emerging presence [1] [2]. The five original stages reported in the 2010 report included the following: The first stage named the *emerging* stage, where governments have established their presence on the web. The second stage is named the *enhanced* stage, where governments started to provide full information about their services and government agencies. The Third stage is the *interactive* stage, where governments started their dialogue with their citizens (and businesses). The fourth stage is labeled the *transactional* stage, where services were conducted and some to full capacity. Finally, the fifth stage was named the *connected* stage, where one interface serves all types of services and departments.

The only downside of such shift is the negligence of the capacity, density, and adequacy of information within each stage. The notion of open government is closely related to information status, where more information is published and at some stage, the quality of information is an indicator of such openness.

The importance of e-government projects is not deniable. More countries across the world are embracing such phenomenon based on two major factors: the first is related to the direction towards more utilization of technology and Internet applications. Even social media started to play a crucial role in defining the relationship between governments and their citizens (or businesses). The second factor is related to the political direction of e-government projects, where some researchers are still considering e-government a direction for improving services [3] [4] [5] [6], while others are embedding political dimensions like participation and transparency [7].

Based on the previous introduction, open government is an important concept where it extending from the concept of e-government. It emphasizes the status of information in each stage. Regardless of the country's achievement in e-government projects, their open government status can be assessed. This paper will review the literature to understand and summarize the concept of open government. Experiences from the World are investigated in the following section. Then a proposed framework will be depicted to guide future research.

2. Literature Review

Open government is an extension of the concept of e-government. It is not a sub-dimension of e-government, or a different concept. Rather, open government is e-government with focus on information status. The concept revolves around freeing information for the purpose of more efficient government work and a better and cheaper service for the public [8]. It is important to realize that public information is useful for citizens and thus needs to

be published and accessed freely [9]. The Obama administration defined open government around three dimensions: transparency, participation and collaboration. The initiative came in effect when the Obama administrative opened a website to grant accessibility to Federal Data to the public under their support for the transparency initiative (www.data.gov). A description of the accomplishments of such initiative will be provided under transparency.

The argument behind such a start depends on the core meaning of open government, where being transparent is the major core dimension. If governments need to know the needs of their citizens, then they should open communication with them, which means more participation. Finally, being open is not about informing and receiving feedback, but enforcing such feedback, which puts governments into the collaborative mode with their citizens. Such process is critical to reach the fully open government state.

We can define open government as a direction of e-government where more information is exchanged between governments and their citizens (businesses) in a transparent, participative and collaborative mode. The open government process needs to keep all characteristics of information to the required quality. The following sections will review sample literature related to the three dimensions.

2.1 Transparency

Information dimensions related to public records include the following: accuracy, validity, security, preservation [9], comprehensiveness, relevancy, timely [7], reliable, and of high quality [10]. Transparency was the major initiative by the Obama administration, where a website was opened (www.data.gov) to grant accessibility to Federal data [11]. Since then, the initiative accomplished few directions like better participation by the public, more data availability, more data disclosure, more compliance to accessibility standards, and a review of existing rules pertaining to open data. Such directions were attained through many initiatives and policies. After two years of the US initiative, an empirical study was conducted to see what factors influenced the three pillars of open government based on the perceptions of US public officials; the authors concluded that publishing accurate and needed reports will result in a better transparency [12].

Transparency is defined as the extent to which governments make available data and documents to the public according to their needs, where they later assess governments' actions and hold them accountable for their actions [13]. The relationship between transparency and e-government is conceptually legitimate, where the more governments use ICT, the more they need to be transparent and vice versa. Such argument is supported by Harrison et al. [10] as they asserted that the shift from information use within e-government websites to more transparent systems is directing towards a more political term which is democracy. Other researchers [14] went further to test if e-government and transparency are related empirically, where an associa-

tion test was conducted for global archival data and concluded to a significant correlation between them [14].

Transparency is critical as the more governments open data, the more the chances they drop in the trap of violating privacy issues [15]. Governments need to protect citizens' and businesses' private data by enforcing the necessary security measures. Based on that, is required more emphasis on the training needs of public officials when enforcing transparent measures of data [16].

Based on the previous summary, we can say that transparency can be defined as the open communication between governments and their citizens guarding for the full capacity of information status (dimensions of information: completeness, time related, relevancy, accuracy, quality, and ownership).

2.2 Participation

E-participation involves "*the extension and transformation of participation in societal democratic and consultative processes mediated by information and communication technologies (ICTs), primarily the Internet*" [17]. Participation is attained through more than one force in society; one of them is open government data. It is conceptually assumed that a link between open government and participation growth would be significant.

A proposition by Goble [18] concluded that linked open data will lead to better public participation. The author proclaimed that less than 5% of the available data sets on UK website are considered linked data. Such issue raises concerns about the progress of such initiatives throughout the world. Some researchers [19] related open government with participation as they proclaimed that open government efforts will yield to better participation. The authors cautioned that such efforts might fall short if the purpose is to introduce the complexity of policy formulation to citizens. On the same line, the size of published linked data is largely contributed to theoretical research contributions and not for applicable industrial data suitable for public use [20].

On other research directions, Abu-Shanab and Al-Dalou' [21] extensively summarized the literature related to e-participation and concluded to five levels of participation: e-informing, e-consulting, e-involving, e-collaborating, and e-empowering. Such premise confuses the dimensions of open government as it considers collaborating as part of the participation process. It is still considered by some researchers that e-participation is a major dimension of e-government, which puts open government as a dimension of participation initiatives [7].

Part of the tools reported to have significant influence on participation is using Wikis [12]. Many tools are supporting participation initiatives and specially web 2.0 tools like websites, social media, wikis, blogs and video publishing websites (examples are included in the work of [21].

2.3 Collaboration

Collaboration is the most controversial dimension among the three pillars of open government, as it means the inclusion of two parties in a process to produce something (a report, information, product, or any other item). Such argument might be easy to administer if the collaborating parties are public and capitulate to same source of authority. But when the two parties are different, like the transactions in G2C and G2B, then the final product is disputable and need to be defined with respect to ownership. Based on that if we carefully review the Obama initiative, we can see that it did not mean to open government without any restrictions, but emphasized the protection of privacy and security and guaranteeing the accessibility of disabled under national legal framework [8].

Collaborative effort by government is key in open government and closely related to transparency [12]. The empowerment of people (the ultimate goal of democracy) is reached by more collaboration between people and their government [22].

Collaboration does not mean G2C only, but might include G2B or G2G initiatives, where great synergies are attained through the cooperation towards saving money and reducing costs. The literature related to collaboration emphasized the requirement of specific tools for the success of collaboration and public participation [23].

We can define collaboration as the joint effort to participate in the democratic process. Such definition can be extended to be a sub-dimension of e-participation, a joint effort between governments and citizens (or businesses), a directed effort towards a product or service, and a legal protection of parties involved in the process and their production. Such definition guards for privacy issues and requires specific tools and systems to allow for such collaborative process.

2.4 Reflections from countries of the world

The experiences of different countries of the World in open government present an accurate picture of current practices and useful insight for research directions. The Indian experience [24] reflects a bottom up pressures that reflects the demand of public towards an open government practices. The authors proclaim that the open government partnership failure is forcing the government and parliament to consider some legislation related to transparency of information [24].

China started its efforts towards open government in 2007, forcing agencies to disclose certain types of information and responding positively to citizens' request to do so. The second step was to review all agencies performance through an annual report (Government Website Performance Review). Finally, measures of performance were established with each constitutions of sub-dimensions, where all indicators were related to open government direction [25].

Other examples were summarized by Mergel [26] and included the UK and Australian catalogues of contracts and spending data (through a dedicated website), the New Zealand portal with public non-

personal data, and the Kenya Open Data Initiative (KODI) that focused on spending information. It is interesting to mention that some researchers doubt the real influence of the Obama initiative in the USA when related to transparency. Open government was described, like transparency, as a double sword strategic step which might result in unrealistic expectations based on full/partial information disclosure [27].

The UK website (data.gov.uk) is a good example of open government initiative, where 5400 data sets are available for use and reuse by diverse stakeholders. The strength of such website is collecting all these data sets in a searchable fashion so stakeholders use it for decision making and more insights. Another good example is the World Bank's website (data.worldbank.org), where country data is organized, re-arranged to be more suitable for navigation, research and decision making [28].

A study of the Spanish laws [29] related to transparency showed that regulations fall short in guaranteeing transparency and citizens' right to information. The author emphasized the contribution of the Spanish E-government Act to citizens' right to information, the accessibility and security of information, and the electronic tools contributing to the diffusion of information like electronic offices [29].

2.5 Summary and Proposed Conceptual Model

Open government initiatives are assumed to enhance the relationship between governments and their citizens [30], and maybe businesses through transparency, participation and collaboration. Accountability is also much enhanced through linked data initiatives [18], which comes as an integral part of open government. One of the reported challenges is the restrictions forced by existing laws, where governments need to resolve legal constraints on data disclosure [15]. It is important to realize the movement towards more laws related to open government, where 93 countries have established some laws or regulations (by the end of 2012 and as stated in [25, p. 4].

Governments spend money on data, where sacrificing such cost might be a challenge for governments. The case proposed by Good speed [15] in relation to GIS data is an example of such obstacle. The author emphasizes the importance of dealing with the practical, ethical, and legal considerations related to accessing public information. Empirical Research in the USA [12] revealed an acceptable satisfaction with the achievements of Federal government in the country. Government officials felt that open government initiatives improved performance and extended government outreach to citizens [12].

Open government can be redefined to be *the actions and policies governments take to make their data and information open based on citizens' needs (and even businesses) through a connected communication, participation and collaboration mode*. The previous few sections did not add much to the open government dimensions. The literature revolves around transparency, participation, and col-

laboration, where segregating the dimensions of the known open government concept can be difficult at this stage. This paper tried to explore the concept and try to understand how it contributes to our knowledge of e-government initiative. The first two stages of government revolve around information, where transparency is crucial. The third evolution stage described earlier revolves around the communication and transaction, where the participation of citizens and businesses is important. Finally, when we reach a connected stage (final one) we need to collaborate as one body to enhance services and improve the decision making process. Based on that, we can summarize our understanding of the open government environment and translate such definition into a conceptual model that is depicted in Figure 1. The model adopt a process with flow direction, where collaboration can be more valuable, but difficult with respect to legal and administrative requirements and infrastructure.

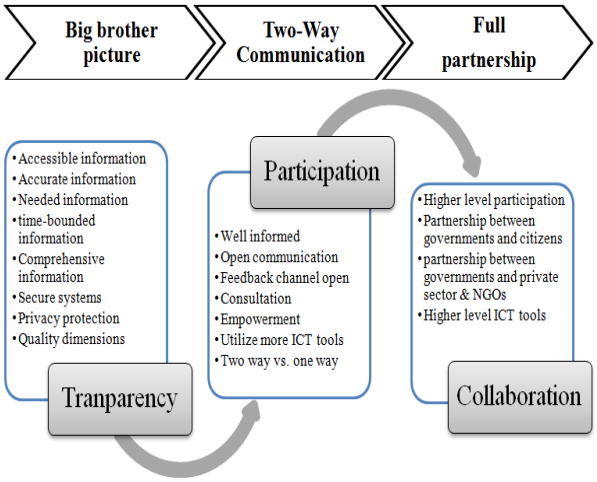


Figure 1: Open government flow model

Developed by the author

The model starts with a focus on transparency, where information flow from government is crucial to the success of e-government. Such perspective can be characterized by being a big brother picture more than further stages. The model (Figure 1) lists the characteristics of such stage in the process.

The second stage focuses on participation, where we try to differentiate between the more developed stages of e-participation in the literature to come up with a focus on two-way communication and transaction of services. The say at this stage is still to governments, but the contribution of citizens is evident. Figure 1 again depicts the characteristics of this stage, where the theme of the process is switched to a two-way communication.

Finally, the model tries to accommodate a position of more developed e-participation stages and segregated the collaboration stage to fit with the known open government initiative. The model depicts a picture of more collaboration and involvement in decision making process where citizens and government are perceived as partners.

3. Conclusion and Future Work

Open government initiatives are limited to three major dimensions: participation, transparency and collaborations. Research [12] indicated that satisfaction with public achievements based on open government is closely related to advancements on the three dimensions [12].

On the other hand, open government initiative is a strategic direction that might not be suitable for all

countries, being open (like open source) is the opposite of intellectual property and copyright protection [31]. Before advancing to open government, some questions need to be answered [32]: To whom information must be open? Which agency should evaluate availability? How to handle copyright or fair use issues?

The debate around open government entrenches into more than one direction, where public information is collected, organized, and stored by governments, but not fully owned and not freely disseminated and reused. The legal constraints of ownership reuse rights, and legitimate representation of real owners (citizens and businesses) is a debatable issue [33], and vulnerable to private sector forces and interests [34]. A study related to the Chinese experience [35] in open government concluded to a major challenge facing the success of such step, which is the attitude of public officials [35]. Research focused more on typologies of e-government frameworks where much research focused on a trilogy of infrastructure, social, and governmental relationships [36]. Finally, issues of trust in e-government become more potent when governments go open [37].

Future work is required to explore perceptions of citizens, business owners, and public officials in relation to the model depicted and its dimensions. Also, an itemized instrument needs to be built to better understand all the dimensions of open government. Such instrument will help researchers explore the domain and investigate its direction. Finally, much jargon revolves around open government efficiency and strategic direction, researchers around the world are invited to test such argument and see if such phenomenon is a fashion or a fad.

4. References

[1]UNDESA (2010). United Nations e-Government Survey 2010: Leveraging e-government at a time of financial and economic crisis. Published by the Department of Economic and Social Affairs, United Nations, New York, 2010.

[2]UNDESA (2012). UN Global E-Government Survey Report 2012, E-Government for people, a report published by Department of Economic and Social Affairs, United Nations, 2012.

[3]Layne, K. & Lee, J. (2001). “Developing Fully Functional E-Government: A Four Stage Model”. Government Information Quarterly, 18, pp. 122-136.

[4]Basu, S. (2004). “E-Government and Developing Countries: An Overview”. International Review of Law, Computers & Technology, 18(1), 2004, pp 109-132.

[5]Evans, D. & Yen, D. (2006). “E-Government: Evolving Relationship of Citizens and Government, Domestic, and International Development”. Government Information quarterly, 23, pp. 207-235.

[6]Yildiz, M. (2007). E-government Research: Reviewing the Literature, Limitations, and Ways

Forward. Government Information Quarterly, 24(2007), pp 646-665.

[7] Abu-Shanab, E. (2013a). Electronic Government, a tool for good governance and better service. A book published by author (deposit number: 2013/2/355, call number: 658.4038), 2013, 195 pages.

[8] Updegrove, A. (2009). How Open Must an Open Government Platform be? A memorandum for the Heads of Executive Departments and Agencies, Subject: Transparency and Open Government, accessed from the Internet in March 2014 from: http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/.

[9] Dawes, S. (2010). Stewardship and Usefulness: Policy Principles for Information-Based Transparency. Government Information Quarterly, Vol. 27(4), pp. 377-383.

[10] Harrison et al. (2012). Open government and e-government: Democratic challenges from a public value perspective. Information Polity, Vol. 17 (2012), pp. 83–97.

[11] Anonymous (2013). United States Open Government Action Plan. Editorial content in Journal of E-Governance, Vol. 36(2013), pp. 71-72.

[12] Reddick, C. & Ganapati, S. (2011). Open Government Achievement and Satisfaction in US Federal Agencies: Survey Evidence for the Three Pillars. Journal of E-governance, Vol. 34(2011), pp. 193-202.

[13] Florini, A. (2007). The Right to Know: Transparency for an Open World, Columbia University Press, 2007.

[14] Abu-Shanab, E. (2013b). The Relationship between Transparency and E-government: An Empirical Support. IFIP e-government conference 2013 (EGOV 2013), September 16-19, 2013, Koblenz, Germany, pp. 84-91.

[15] Goodspeed, R. (2011). From Public Records to Open Government: Access to Massachusetts Municipal Geographic Data. URISA Journal, Vol. 23(2), 2011, pp. 21-32.

[16] Kimball, M. (2011). Mandated state-level open government training programs. Government Information Quarterly, Vol. 28(2011), pp. 474–483

[17] Saebo, O., Rose, J., & Flak, L. S. (2008). The shape of eParticipation: Characterizing an emerging research area. Government Information Quarterly, 25(3), 400-428.

[18] Goble, C. (2013). Linked Data in Government. IEEE Internet Computing, Published by IEEE Computer Society, July/August 2013.

[19] Evans, A. & Campos, A. (2013). Open Government Initiatives: Challenges Of Citizen Participation. Journal of Policy Analysis and Management, Vol. 32(1), pp. 172–203.

[20] Rodríguez-Gonzalez, A., Valencia-Garcia, R. & Colomo-Palacios, R. (2012). Guest editors

introduction to the International Journal of Software Engineering and Knowledge Engineering, Vol. 22(3), 2012, pp. 323-324.

[21] Abu-Shanab E. & Al-Dalou', R. (2012). E-participation Initiatives: A Framework for Technical Tools. International Arab Conference of e-Technology (IACe-T'2012), pp. 57-64.

[22] Al-Dalou', R. & Abu-Shanab, E. (2013). E-Participation Levels and Technologies. The 6th International Conference on Information Technology (ICIT 2013), 8-10 May, 2013, Amman, Jordan, pp.1-8.

[23] Luna-Reyes, L. & Chun, S. (2012). Open government and public participation: Issues and challenges in creating public value. Information Polity, Vol. 17(2012), pp. 77-81.

[24] Dey, N. & Roy, A. (2013). India in Open Government and Open Government in India. Stanford Social Innovation Review, accessed from the Internet in April 2013, from: http://www.ssireview.org/articles/entry/india_in_open_government_and_open_government_in_india.

[25] Zhang, J. & Li, C. (2013). China's Open Government Efforts: Progress and Challenges. PA Times, Vol. 36(1), pp. 4 & 13.

[26] Mergel, I. (2012). Open Data Goes Global: Challenges and Solutions. PA Times, Vol. 35 (1), pp. 7-8.

[27] Coglianese, C. (2009). The Transparency President? The Obama Administration and Open Government. Governance: An International Journal of Policy, Administration, and Institutions, Vol. 22(4), October 2009, pp. 529–544.

[28] UNDESA (2013). Guidelines on Open Government Data for Citizen Engagement, a report published by Department of Economic and Social Affairs, United Nations, 2013.

[29] Cerrillo-i-Martínez, A. (2011). The regulation of diffusion of public sector information via electronic means: Lessons from the Spanish regulation. Government Information Quarterly, Vol. 28 (2011), pp. 188–199.

[30] Unsworth, K. & Townes, A. (2012). Transparency, Participation, Cooperation: A Case Study Evaluating Twitter as a Social Media Interaction Tool in the US Open Government Initiative. The Proceedings of the 13th Annual International Conference on Digital Government Research, June 04 - 07 2012, College Park, MD, USA, pp. 90-96.

[31] Tkacz, N. (2012). From open source to open government: A critique of open politics. Ephemera articles, theory & politics in organization, Vol. 12(4), pp. 386-405.

[32] Bremer, E. (2012). Incorporation By Reference; In An Open-Government Age. Harvard Journal of Law & Public Policy, (based on a government report), Vol. 36(1), pp. 131-210.

[33] Henninger, M. (2013). The Value and Challenges of Public Sector Information. Cosmopolitan Civil Societies Journal, Vol. 5(3), 2013, pp. 75-95.

[34] Bates, J. (2012). This is what modern deregulation looks like: co-optation and contestation in the shaping of the UK's Open Government Data Initiative. The Journal of Community Informatics, Vol. 8(2), Accessed from the Internet in April 2014 from: http://ci-journal.net /index.php /ciej/article/view/845.

[35] Piotrowski, S., Zhang, Y., Lin, W. & Yu, W. (2009). Key Issues for Implementation of Chinese Open Government Information Regulations. Public Administration Review, Vol. 69(1-December 2009), pp. 129-135.

[36] Abu-Shanab, E. (2012). Digital Government Adoption in Jordan: An Environmental Model. The International Arab Journal of e-Technology (IAJeT), Vol. 2(3) January 2012, pp. 129-135.

[37] Abu-Shanab, E. & Al-Azzam, A. (2012). Trust Dimensions and the adoption of E-government in Jordan. International Journal of Information Communication Technologies and Human Development, Vol. 4(1), 2012, January-March, pp.39-51.

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of Information Technology and Networking

Article

BANKING E-SERVICES REQUIREMENTS IN YEMEN

1.Dr. Arwa Al-eryani *, 2.Samah Alaraki
1.Yemen Academy for Graduate Studies, Yemen , 2.Yemeni University, Yemen

Article info

Article history:

Accepted Jan, 2015

Keywords:

E-services,
Requirements,
Banking.

Abstract

E-banking is one of the most popular services utilized by most of the country's banking recently. Studies have shown that the quality of e-services is a core issue that influences adoption of e-banking services worldwide. The objective of this study is to investigate the availability of the essential requirements for e-banking adoption in Yemen. Results show that the infrastructure requirements are the strongest points, where the requirements related to employees and clients are the weakest points related to applying e-service in Yemeni banking. The implications are discussed and recommendations are made in order to improve e- banking service in Yemen.

* Corresponding author: 1.Dr. Arwa Al-eryani,
E-mail address: Arwa_y@hotmail.com

Introduction

In the past decade there were several discussions around those new terms; e-Management, e-Business and e-Commerce. They were marketed as the magic solution to all companies' problems. As any new technology they mostly remain only at a theoretical stage, and were implemented first in large companies [1] .The shift of societies towards the information society has had deep effects on numerous aspects of human life such as economic, social and cultural aspects [2]. The impact of information technology on human societies is not less than that of industrial revolution; therefore the information technology development and its applications are regarded as the fourth digital revolution [3].

As information technology grows, e-business applications are found all over the world. More and more companies recognize the benefits of e-business and move from traditional business to e-business [4]. E-management is the modern way to manage all e-business application [3]. It can also play a role in e-business and e-commerce [5]. The barriers that effective e-business are varied enough in the industrial world, but especially problematic in the developing world [3]. We use the term "e-management" to broadly describe the publishing of information and the performing of various transactions and services over the Internet, Extranets, or Intranets. It seems that the Internet and e-business have changed the way firms conduct business globally using e-management [6].

The implementation of e-management is related to all the employees of the company. That is why we have to consider the human factor as the most important factor impacting e-management [7]. The human factor plays a big role in e-management success and security [8]. Understanding how the Web fits into a country's culture is necessary to form successful customer relations [3].

Banking has also tried to benefit from the new environment to offer quality services and gain new customers. Online banking allows the customers to perform transactions, pay bills and check balances 24 hours a day, 7 days a week. The bank virtually never closes because it is accessible from any PC or laptop computer [9]. Through the Internet, transactions are typically performed and executed at a faster rate than traditional method. In the other hand, one of the biggest barrier which faces using e-transaction is the trust [8]. The customers may wonder if their transactions went through successfully or if they clicked on the correct button or not.

Research theoretical

E-management

E-business is the administration of conducting business via the Internet [10]. E-management will be critical for ensuring the availability of e-business applications anywhere and anytime [11]. The implementation of e-management is related to all the employees and customers of the company. That is why the human factor is considered as the most important factor impacting e-management [7], [11]. The culture and sensitivity to cultural differences plays a critical role in successful in-

ternational business, and just as much as for international e-business [3]. The most benefits from using e-management is the elimination of distances through providing linkage among separate computers in the Internet, and the computerization of systems and telecommunication which result in new capacities to transfer sounds and images [11]. Also it ensures the best utilization of resources, increases efficiency, and provides support to high management in planning, and managing human and financial resources [12]. E-management can play a role in e-business & e-commerce management, and gain some benefits such as[5]:- reducing costs, improving product or service quality, reaching new customers or suppliers, creating new ways of selling or providing existing products and services. The issues of security of transactions and privacy protection over the Internet are of concern to many consumers [6]. E-service is a part of e-management, it means services, action or set of actions of either a commercial or non-commercial nature, and includes the provision of information and/or e-government services [13].

E-banking

E-banking is the electronic bank that provides the financial service for the individual client by means of Internet [14]. Meaning that, any user with a personal computer and a browser can get connected to his bank's website to perform any of the virtual banking functions. To make a step towards e-banking, a bank must commit itself to many substantial strategy changes such as it must allow the customers to participate more actively in the carrying out of the bank service [9]. There are many functions that can be offered by e-banking. At present, the personal e-bank system provides many service, like; Inquiry about the information account, Card account 'transfer, Bank-Securities accounts, Account management [14].

Yemen banking

The Cooperative and Agricultural Credit Bank (*CAC Bank*) has offered a *CAC Online* to meet the customers banking wishes and needs and facilitates their transactions via the internet. Through *CAC Online* service, the customers have access to and could review their accounts at any time and from any place. *CAC Online* provides: Confidentiality of data and banking operations according to the highest degree of security and advanced technology [15].

Yemen Commercial Bank (*YCB*) is a unique and modern banking launch represented in the executing comprehensive banking transactions and applying the electronic and prompt Network services among branches which made a distinctive transmission in the level of efficiency and performance of bank services. YCB was able also to utilize the modern technological progress to provide electronic banking services including ATM, Sale points, phone banking, SMS and money gram transfers and Speed Cash service [16].

Arab Bank has offered a convenient and secure way to manage the customer accounts, anytime and anywhere. The e-services includes; inquiring about the balance, viewing accounts summary, viewing, saving, printing account statement & transaction details, accessing cards to review details, paying credit card and viewing card transac-

tions, viewing fixed deposit details, payinf utility bills, transferring funds between accounts, transferring funds to other Arab Bank customer within the same country, transferring funds to other accounts worldwide, changing your address, sending instructions or inquiries through secure email, setting up scheduled payment order [17].

Yemen Bank for Reconstruction and Development has started using technological progress to satisfy the needs of their valued customers. They extended a collection of distinct services round the clock (24 hours) during the week days [18].

Tadhamon International Islamic Bank (*TIIB*) provides its customers with new developments in banking services and offers the services to customers. *TIIB Online Banking* allows their customers to keep track of their finances in a safe and secure environment, 24 hours a day, 7 days a week [19].

Shamil Bank of Yemen & Bahrain has launched Internet Banking (*Shamil Online*) at the beginning of 2013. This service enables a bank customer to examine the details of his account and performs many bank transactions through the internet web, without visiting the Bank premises. It also enables the customer to manage his/her accounts in privacy and safety [20].

Saba Islamic Bank provides their customer with many services through their site, such as keeping track of their finances in a safe and secure environment [21].

Literature Review

Rahman *et. al.* [11] indicated that the e-management is an umbrella name for several e-business modules. Some barriers have addressed which have impact in e-management implementation in Iran:- (1) Managerial Factors, e.g. lack of technological awareness among managers, lack of awareness among the mangers about the advantages of IT; (2) Humanistic Factors e.g. lack of IT specialists in organizations, employees` lack of interest and motivation to apply new techniques; (3) Cultural-Social Factors e.g. the non-developed culture for proper application of IT, unfamiliarity of users with IT and unfamiliarity of citizens and authorities with IT performance; (4) Organizational-Structural Factors e.g. weakness of communication channels in organizations, and lack of financial resources; (5) Technical-Technological Factors e.g. lack of sufficient software facilities, incongruity between systems and users; (6) Environmental Factors: e.g. no integrated network in country, lack of necessary rules and regulations in country.

Some of the e-management challenges which have been addressed are expanded competition, increased customer power; network security has developed, reliability of website technologies and network uptime and speed [11]. In Yemen the implementation of e-management in governmental and private organizations in the city of Aden has been addressed [22]. The results showed that there is awareness at the governmental organizations to offer the management and human recourses needed for the implementation of e-management. Furthermore, good understanding of the importance

offering the security measures needed to implement the e-management was updated [22].

Sikdar *et. al.* [23] have studied the e-banking in India. They introduced the concept of e-banking and its benefits provides from bank's customers as well as banking organization's point of view. The Commercial banking operations in the long-run can be optimized by minimizing the branch based interactions with the customers, enabling the banks to focus on direct selling and other activities requiring higher focus on the part of commercial banks. e-banking as a tool can be used in two objectives.

First bank can optimize its branches operations by doing away with large volumes of routine transactions at individual branches. Second they can expand their geographical presence by reaching the interior parts of the country hitherto cut-off from the ambit of banking services. This will also help in achieving of inclusive banking, strongly propagated by the national banking regulator Reserve Bank of India (RBI).

Jun *et. al.* [24] have studied the issues associated with e- banking service quality. They identified a total of 17 dimensions of e- banking service quality, which can be classified into three broad categories customer service quality, banking service product quality, and online systems quality. Every classified category has many dimensions. Customer service quality dimensions are reliability, responsiveness, competence, courtesy, credibility, access, communication, understanding the customer, collaboration, and continuous improvement.

Online systems quality dimension are content, accuracy, ease of use, timeliness, aesthetics, and security. Banking service product quality has one dimension of product variety/diverse features.

Ta-Wei *et. al.* [Ta 25] investigated the relationships among brand equity, loyalty, trust, satisfaction, and e-service quality in e-Banking. The significant simultaneous relationships between satisfaction and loyalty and between trust and loyalty are confirmed in this study, implying that these relationships must be determined simultaneously, rather than sequentially.

Usman *et. al.* [26] studied e-Banking security, they pointed out that fraud is an issue being experienced globally and is continuing to prove costly to both banks and customers. Frauds in e-banking services occur as a result of various compromises in security ranging from weak authentication systems to insufficient internal controls. Security issues are major barriers to internet banking and e-commerce activities among consumers. This is achieved by understanding business goals, objectives and critical success factors when planning the security strategy, as well as the impact on the business if they are not achieved [Usman].

Fragata *et. al.* [27] proposed two models for the determinants of e-banking loyalty for large business customers. The results demonstrated that five main quality dimensions were identified for the e-

banking portals: assurance, reliability, convenience and quality monitoring by the financial director of the company.

These results also confirmed that both switching costs as well as e-banking quality have a strong impact on e-banking loyalty. The last is being achieved via the mediating effect of e-trust [27].

Aim

This research attempts to investigate the availability of the Banking e-services requirements in Yemen. We investigated the availability of the factors that are critical in the adoption of e-services in banking. This study make a new contribution as it is one of the first researches in Yemen to highlight and draw attention to the situation of the e-services in Yemen banking.

Research Methodology

From the literature studies we can set five main factors to verify the availability of the essential requirements for implementation of e-services in banking including: (1) *Managerial Factor:* to measure the availability of the support of the top management and their awareness.

(2) *Management Factor:* to measure the availability of the e-services processes and procedures. (3) *Technical infrastructure Factor:* to measure the availability of the well-built technical infrastructure required for e-services. (4) *Employment Factor*: to measure the availability of skilled employees required for the new environment and their motivation. Finally (5) *Clients Factors:* to measure the availability of their interest and ability to communicate with the e-services.

The methodology of the research is based on a questionnaire to gather the required information from the top management and managers of the e-services department in each bank (see table 1,2,3,4,5).

The questionnaires have been distributed by email for each key person (top-manger and e-services manager) in each bank. When the questionnaires were returned, they were analyzed and the results were presented.

Results and Discussion

After analyzing the results of the distributed questionnaires we came out with six figures that present the results. We used tables to clarify the statements under each factor.

"*The requirements that are supported by the top management*" is the first factor. It aims to gather information about the following statements: top management is aware of the importance of the e-services in supporting the bank's work, top management understands the e-services implementation barriers, top management works on overcoming the e-services implementation barriers, top management encourages employees to make their own decisions in their departments, top management cares about offering development training for its employees locally, top management cares about offering development training for its employees abroad, top management cares about setting the balance for e-services (see table(1)).

The requirement that are supported by top management factor	
A	Top management is aware of the importance of the e-services in supporting the bank's work
B	Top management understands the e-services implementation barriers
C	Top management works on overcoming the e-services implementation barriers
D	Top management encourages employees to make their own decisions in their departments
E	Top management cares about offering development training for its employees locally
F	Top management cares about offering development training for its employees abroad
G	Top management cares about setting the balance for e-services

Table (1): The requirements that are supported by top management factor

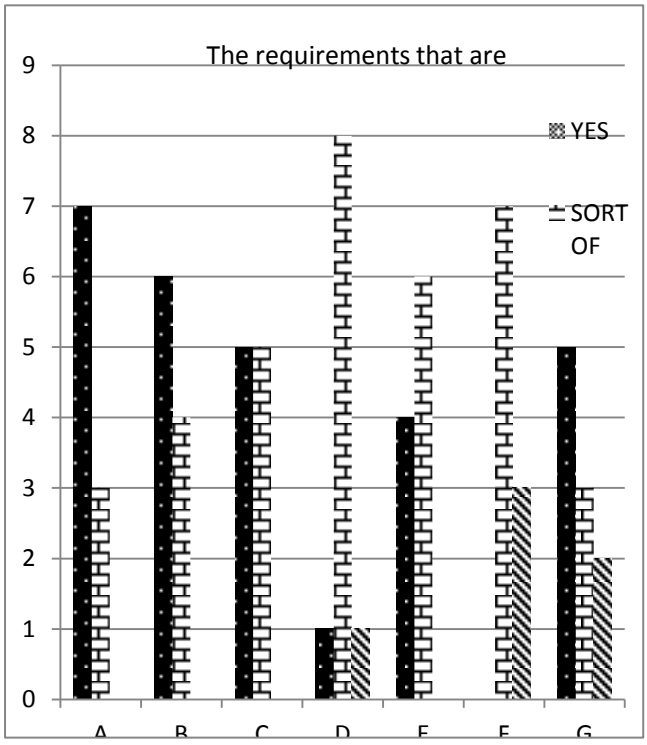


Figure (1): The requirements that are supported by top management factor

importance of the e-services in supporting the bank's work as well as of the barriers, although it did not always overcome them, where we found that 50% of respondents indicated that.

From our previous studies, the importance to encourage the employees to take the decision at their departments was recommended, as this makes them more understandable with the new services. This point was one of the weakness points for this factor. 10% only from respondents indicated that the employees have been involved in decision taking. Also, top management does not care about offering development training for its employees abroad. The balance for e-services was one of the problem facing e-services.

"*The managerial requirements to support e-services*" is the second factor. It aims to gather information about the following statements: the e-services procedures are clear for the client, the bank has a specialized department for developing and monitoring the e-services, the e-services procedures are clear for the employee, the procedures manual is used to execute the e-services, the bank complies with the set times to execute the e-transaction, the bank complies the awareness of the importance of the e-services, and the bank complies to separate the e-culture among the client and employment(see table(2)).

Table (1) and Figure (1) show that the strength point was that the top management is aware of the

The managerial requirements to support e-services factor	
A	The e-services procedures are clear for the client
B	The bank has a specialized department for developing and monitoring the e-services
C	The e-services procedures are clear for the employee
D	The procedures manual is used to execute the e-services
E	The bank complies with the set times to execute the e-transaction
F	The bank complies the awareness of the importance of the e-services
G	The bank complies to separate the e-culture among the client and employment

Table (2): The managerial requirements to support e-services factor

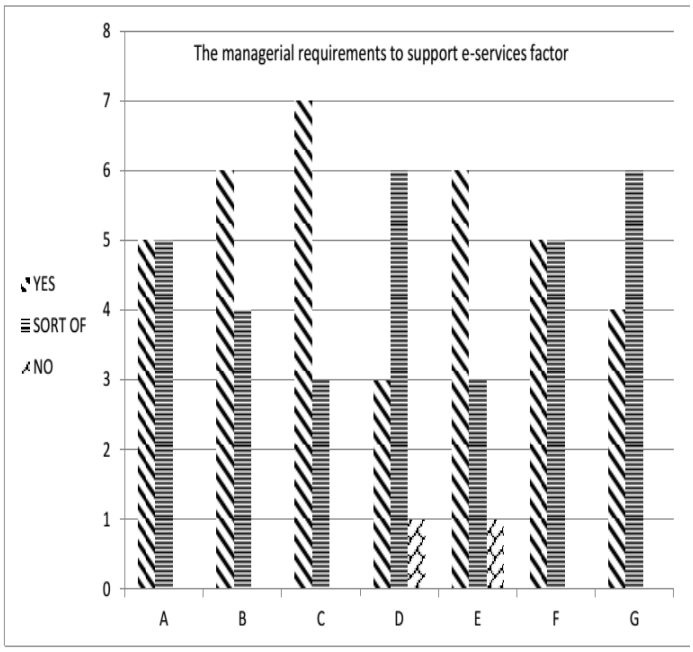


Figure (2): The managerial requirements to support e-services factor

Table (2) and Figure (2) show that the strength points are that all banks have specialized department for developing and monitoring the e-services, The e-services procedures are clear for the employee, the bank complies with the set times to execute the e-transaction. Although most of the respondents agreed that the e-services procedures are clear for the employee, they pointed out that it is not the same way for the clients. The weakness points in this factor were that the bank did not make the announcements of the importance of the e-services, nor separate the e-culture among the clients and employees.

"*Infrastructure requirements*" is the third factor. It aims to gather information about the following statements: the bank has powerful database systems, the bank has powerful security programs to protect its systems, the bank has a dynamic website on the Internet, the bank has systems on its website allowing clients to interact with it, the bank has security systems to protect its transactions over the Internet, the bank has good maintenance to support its networking and software, the bank offers various training courses to develop its employees in using modern technologies, there are clear laws for the e-services, and the bank works on awareness for clients through showing the importance of the e-services(see table(3)).

The infrastructure requirements factor	
A	The bank has powerful database systems
B	The bank has powerful security programs to protect its systems
C	The bank has a dynamic website on the Internet
D	The bank has systems on its website allowing clients to interact with it
E	The bank has security systems to protect its transactions over the Internet
F	The bank has good maintenance to support its networking and software
G	The bank offers various training courses to develop its employees in using modern technologies
H	There are clear laws for the e-services
I	The bank works on awareness for clients through showing the importance of the e-services

Table (3): The infrastructure requirements factor

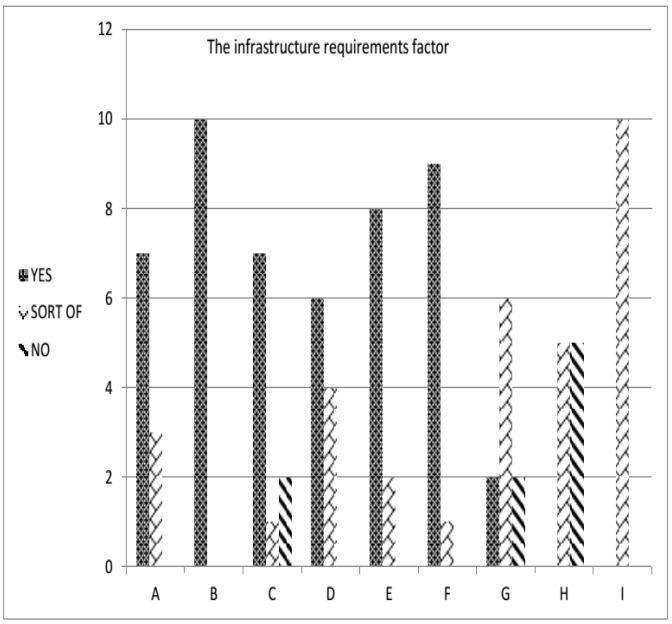


Figure (3): The infrastructure requirements factor

Table (3) and Figure (3) show a positive side for the e-services at Yemen bank, all respondents agreed that their banks have powerful databases systems and powerful security programs to protect their systems; also it has good maintenance systems to support networking and software. The weakest point in this factor was the lack of clear laws for the e-services. The absence of the clear laws and legal rules affected the e-services in great deal where no respondent indicated the exits of these laws and rules.

"*Employees*" is the fourth factor. It aims to gather information about the following statements: the employees have enthusiasm to work through implementing e-services, the employees have the skills to use a computer in a good way, the employees have the skills to work on the Internet, the employees are encouraged to participate in the decision making process, the employees have full comprehension of the importance of the e-services, the employees have the skill to use the hardware and software, and the employees have high salaries(see table(4)).

The Employees factor	
A	The employees have enthusiasm to work through implementing e-services
B	The employees have the skills to use a computer in a good way
C	The employees have the skills to work on the Internet
D	The employees are encouraged to participate in the decision making process
E	The employees have full comprehension of the importance of the e-services
F	The employees have the skill to use the hardware and software
G	The employees have high salaries

Table (4): The Employees factor

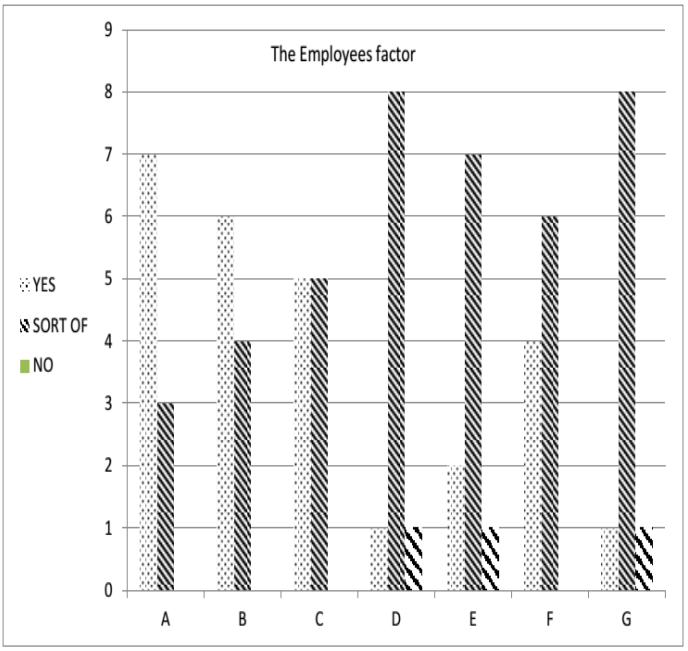


Figure (4): The Employees factor

Again the humanistic aspect is still the weakness point; table (4) and Figure (4) show that the employees don't have full comprehension of the importance of the e-services to understand the importance of the e-services. All respondents indicates that employees have good skills to use computer but they do not have enough skills to use Internet, also the employees do not participate in decision making and this point makes them far from the reasons and benefits of the new developing. When we came to the economic aspect, we can see that 80% of the respondents pointed that the employees do not get good salaries.

"*Clients*" is the fifth factor. It aims to gathering information about the following statements: the clients interact with the bank website, the clients show trust in the bank systems on the Internet, the clients have proper skills to use the e-services, the clients find quick responses to their inquires in general, the clients find quick responses to the problems they face during electronic transactions, the clients can follow the progress of their transactions through the bank website, the e-services is executed faster than the traditional way(see table(5)).

From our previous studies we can understand the importance of clients for success of e-services. They are the only part which makes the e-services useful and gets the benefits.

The Clients factor	
A	The clients interact with the bank website
B	The clients show trust in the bank systems on the Internet
C	The clients have proper skills to use the e-services
D	The clients find quick responses to their inquires in general
E	The clients find quick responses to the problems they face during electronic transactions
F	The clients can follow the progress of their transactions through the bank website
G	The e-services is executed faster than the traditional way

Table (5): The Clients factor

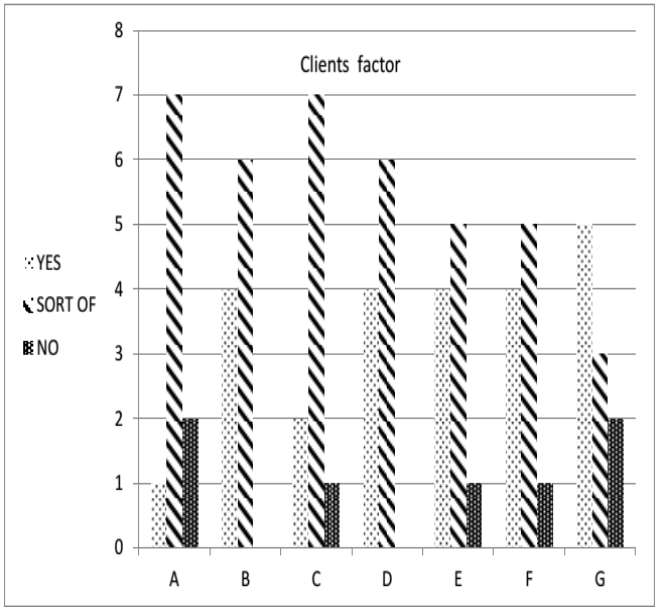


Figure (5): The Clients factor

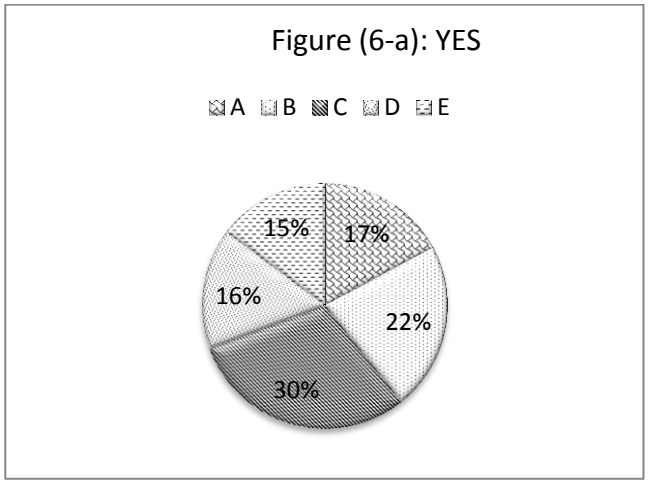
Table (5) and Figure (5) show that there is a limited interaction between clients and the bank e-services. Many reasons might be behind this, first and as shown from the respondents answers the clients do not get quick responses to the problems they face during electronic transactions, the clients can not follow the progress of their transactions easily, and the trust is still a main problem between the clients and the e-services.

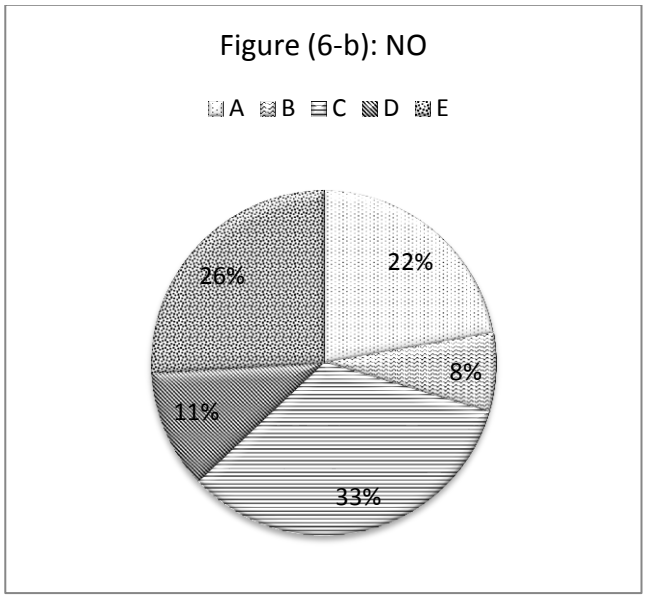
Conclusion

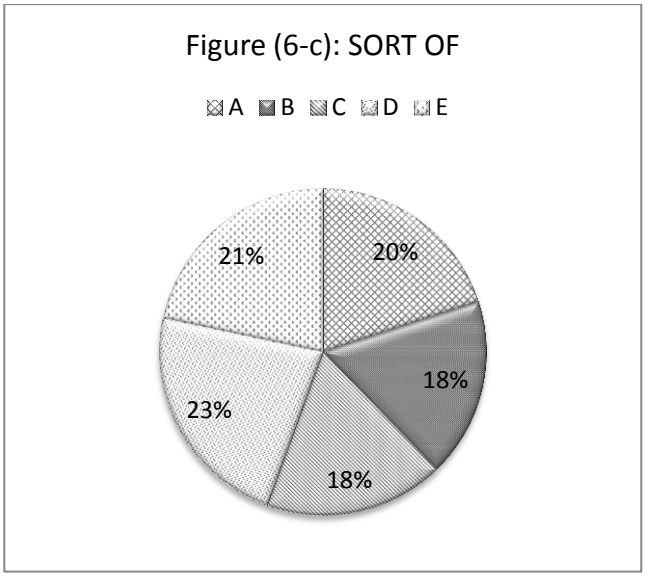
From table (6) and figures (6-a, 6-b, 6-c) which presented our five main factors; The requirements that supported by top management factor, The managerial requirements to support e-services factor, The infrastructure requirements factor, The Employees factor, The Clients factor. We can come up to the following results; the Yemen banks have good infrastructure requirements, good top management supporting especially in the understanding of the importance of e-services.

A	The requirements that are supported by top management factor
B	The managerial requirements to support e-services factor
C	The infrastructure requirements factor
D	The employees factor
E	The clients factor

Table (6): Factors Information







On the other hand the main problems were; the e-services laws and rules have not been established completely and clearly. The human aspect and e-culture still impact the success of e-services; employees do not have good training courses, they do not have good salaries to encourage them to work with e-services. The clients' attention isn't drawn to the importance of the e-services and the trust isn't built strongly between the clients and the e-services.

Recommendation

Yemen as one of the developing countries face a lot of economic problems which are always consider as big barriers when apply the new technology as e-services. From our research results and analysis we can come up with the following recommendation:-

- The e-services laws and protection rules should be set and announced to build the trust and confidence to use e-services.
- Paying attention to qualify the employees and offer local and abroad training courses.
- Building new relationship with employees by giving them more shares in the decisions taking and respects their suggestion.
- Building strong trust between clients and the e-services by the applying a quick response to their queries and problems they face during their use of e-services and give them the methods and ways to follow their transactions progress.

References

[1] Iulian, M., (2008) "Implementing e-management in small and medium enterprises", available via http://steconomiceuoradea.ro/anale/volume/2008/v4-management-marketing/259.pdf

[2] Dibrell, C.; Miller, T. (2002.) "Organization design: The continuing influence of information technology", available via http://www.emeraldinsight.com.

[3] Granville, B.; Leonard, C., & Manning, J. (2001), "Information technology and developing countries: potential and obstacles" available via htt://www.tcf.or.jp/tcp/tcp14/RIIA-information.pdf .

[4] Jingting, L. ; Jinghua, H., (2004), "An Exploratory Study of E-Business Success Factors", Journal of Electronic Science and Technology of China Vol.2 No.3.

[5] Chaffey, D. (2011), "eBusiness & eCommerce Management", Prentice Hall.

[6] Patrick C.; Peter R., "Barriers To Effective E-Business In Developing Countries", International Business & Economics Research Journal Volume 1, Number 4 39.

[7] IBM. (2002), "Ten Success Factors for E-Business", http://www-4.ibm.com/software/info/soul/st/guide.html.

[8] Al-eryani, A., (2014), "Review Study of e-management Barriers, Challenges and Security", Saba Journal of Information Technology and Networking (SJITN), Yemen

[9] Hultén, S. ; Nyberg, A. and Chetioui, L.(2002) "First-mover advantages and disadvantages in e-commerce – the example of Internet banking" European Integration in Swedish Economic Research conference-Mölle.

[10] Huff, L.; Wade, M. (2000), "Cases in Electronic Commerce", Boston: Irwin/ McGraw-Hill, available via http://geomart25.files.wordpress.com/2014/05/dave-chaffey-e-business-and-e-commerce-management-strategies-4th-ed-qwerty80.pdf.

[11] Rahman, H.; Fayyazi. M. and Asl. N.,(2008), "E-management: Barriers and challenges in Iran", E-ower Kraklead, available via http://www.g-casa.com/PDF/Krakow%202008/krakow%20papers%20pdf/paper%20database%20krakow/Nastaran.pdf

[12] Al-Malik, B., (2007) "Administrative and Security Dimensions of e-management in Saudi Banks", available via file:///C:/Users/user/Downloads/articles_2007_m_as_19%20(1).pdf.

[13] Establishment of Harmonized Policies for the ICT Market in the ACP, (2012) HIPSSA – SADC Model Law on Electronic Transactions & Electronic Commerce.

[14] Vyas, S. "Impact of E-Banking on Traditional Banking Services, India", availa-

ble via http://arxiv.org/ftp/arxiv/papers/1209/1209.2368.pdf

[15] The Cooperative and Agricultural Credit Bank (CAC Bank) website available via www.cacbank.com.ye

[16] The Yemen Commercial Bank (Tijari) website available via www.ycb.com.ye

[17] Arab Bank website available via www.arabbank.com

[18] Bank for Reconstruction and Development website available via www.ybrdye.comYemen

[19] Tadhamon international Islamic bank website available via www.tiib.com

[20] Shamil Bank of Yemen and Bahrain (an Islamic bank) website available via www.sbyb.net

[21] Saba Islamic Bank website available via http://sababank.net/

[22] Mohsen, F., (2010), "the degree of availability of the requirements to implement e-management in the Yemeni governmental organizations", field study in ministry offices and governmental facilities in Aden, Master thesis, Aden University.

[23] Sikdar, P.; Makkad, M., (2013) "Internet Banking in India –A perspective on benefits and challenges involved", International Journal of Engineering, Business and Enterprise Application

[24] Jun, M. ; Cai, S. (2001) "The key determinants of Internet banking service quality: a content analysis", International Journal of Bank Marketing, International Journal of Bank Marketing, 19/7, 276-291

[25] Ta-Wei (Daniel) Kao; Winston T. Lina and Hsin-Hsin C., (2013), "Managing Brand Equity in e-Banking: A Simultaneous Equation System Approach, Managing Brand Equality in E-Banking", available via file:///C:/Users/user/Downloads/DSI%202013%20E-Banking%20Paper.pdf

[26] Usman, A. ; Shah, M., (2013), "Critical Success Factors for Preventing e-Banking Fraud",
Journal of Internet Banking and Commerce, vol. 18, no.2, An open access Internet journal (http://www.arraydev.com/commerce/jibc/).

[27] Fragata, A. ; Moustakas, E., (2013), "Investigating the Determinants of e-Banking Loyalty for Large Business Customers: Two Empirical Models", Journal of Economics, Business and Management, Vol. 1, No. 2, May 2013

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of
Information Technology
and Networking

Article

DEVELOPMENT OF WIRELESS SENSOR NODE FOR SPORTSMEN

Galal Khaled Al-Dhamari

Don State Technical University, Russia

Article info

Article history:

Accepted Dec, 2014

Keywords:

Wireless,

Sensor,

Node,

Sportsmen

Abstract

In last years, the vital data of sportsmen at the time of exercise and game have been used to plan efficient exercise and elude injuries and diseases, where a vital sensor node is frequently installed at the chest position of a sportsman. One of the major problems is that a tri-axis accelerometer sensor should be installed at a different position such as rear waist for exact energy expenditure (EE) evaluation whereas an electrocardiogram (ECG) sensor should be separately installed at a chest position for exact heart rate (HR) evaluation.

In this paper, afresh Development a wireless waist-mounted vital sensor node for both EE and HR evaluation. First of all, to select a microcontroller applicable for the data processing in the vital sensor node, we computationally compared three microcontrollers in requirements of energy consumption necessary for the vital data processing. The matching showed that LPC1768 is the most power- saving candidate, so then designed and realized prototype vital sensor node utilization LPC1768.

* Corresponding author: Galal Mohammed Al-Dhamari
E-mail address: Jalal_exe@mail.ru

Introduction

Recently, in the field of sports, the vital data of sportsmen at the time of exercise and game have been used for analyzing their physical strength and preventing injury. Wireless vital data collection system [1] has drawn much attention for realizing real-time monitoring of sportsmen' performance, which is helpful for their coach and trainer on the field. Here, important vital information includes heart rate (HR), energy expenditure (EE) and body temperature (BT).

Wan-Young Chung and his teams [4], had designed a non-intrusive healthcare system, This system was designed for a wide area coverage which is based on wireless sensor network to support RF transmission, Through this project, they developed a system that allows transmitting physiologic data in wireless sensor network using IEEE 802.15.4. The data were transmitted to a base station which is connected to a server PC from on-body wearable sensor device, which consist of wearable USN node, wrist pulse oximeter and chest sensor belt.

Tia Gao's team [5] designed their prototype of patient monitoring and tracking systems based on CodeBlue project This prototype consists of a wearable computer that attached to the patient wrist which known as smart dust or a mote which transmit the information to the first responder's tablet device continuously. This project allows data transmission uses the TinyOs Active Messages Protocol which is based on IEEE 802.15.4. To allow indoor and outdoor usage, they implement the system with GPS to provide geolocation and indoor location detection system. This system is portable and can be use everywhere.

Vital sensor nodes are often installed at the chest position of a sportsman because an electrocardiogram (ECG) sensor is used for HR estimation, like SPIHPU from GPSports [2]. One of the main problems is that an ECG sensor should be installed closer to heart for correct HR evaluation, whereas a tri-axis accelerometer sensor should be separately installed at a waist position for accurate EE evaluation [6].

To solve the problem, we have proposed a wireless waist-mounted vital sensor node, in which "HR evaluation using weak ECG signal sensed at waist position" is required. In our previous work, we have shown that such HR evaluation is possible, developing a time-domain signal processing technique for computing HR from the weak ECG signal and demonstrating its performance by an experiment [3].

In this paper, we recently development a waist-mounted wireless sensor node for transmitting vital data sensed from sportsman using the 920 MHz frequency band. First of all, we discuss the signal processing capability necessary for the wireless waist-mounted vital sensor node. Then, we compare the productiveness of three candidate microcontrollers in terms of "power consumption" and "accuracy" for the HR evaluation, and also discuss the trade-off between selection of microcontroller and resultant energy efficiency. After selecting the best microcontroller out of the three candidates, we finally show the specifica-

tions of a prototype vital sensor node using the selected microcontroller.

The rest of the paper is organized as follows. Section II shows the overview of our vital sensor node. Section III shows the requirements for microcontroller in our vital sensor node and compares the three candidates. Section IV shows the design of our prototype vital sensor node. Finally, section V concludes the paper.

Overview of Vital Sensor Node

Fig. 1.a shows the interconnection of measuring time and signal processing capability (performance) required for the vital sensor node. The sensor node detects the ECG signal at a rear waist position far from heart, so the detected signal inevitably becomes very weak, resulting in requirement of a very high-gain amplifier in the analog part and a bandpass filter of higher order in the digital part. Therefore, our proposed sensor node needs a high performance microcontroller. On the other hand, the Holter monitor for the medical purpose should be workable with a small battery for more than 48 hours, and in addition the activity monitor for healthcare purpose needs much longer battery lifetime such as more than several months. However, taking into consideration that exercise and game periods in common sports are characteristically less than two hours, our proposed sensor node can select a high performance microcontroller as long as it can run for only few hours.

Fig. 1.b shows the functional overview of the vital sensor node. The sensed ECG signal, tri-axis acceleration signal and BT signal are input to the sensor node, the HR and EE are correspondingly evaluated by using the ECG signal and the acceleration signal, and finally, the evaluated data are transmitted to the data collection nodes put around a field with transmission power of 20 mW and the data transmission rate of 100 kb/s via a wireless channel in the 920 MHz frequency band. Note that all the functions of the sensor node, such as vital signal detection, processing and transmission, are driven by a small Lithium Ion Polymer battery for light weight.

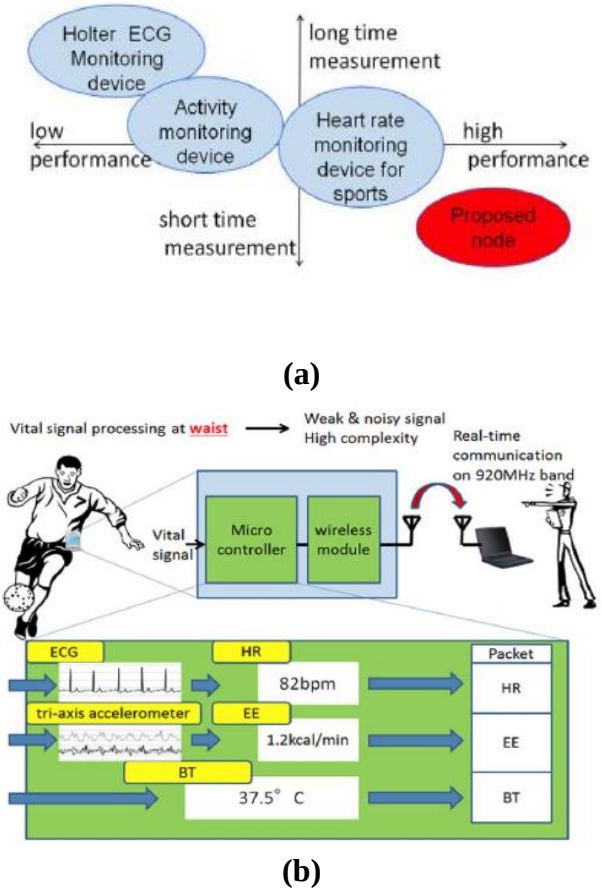


Fig. 1.a Measuring time/signal processing capability for our sensor node, b. Functional overview of the vital sensor node.

Choice of Microcontroller

Matching Of Microcontrollers

The vital sensor node is composed of three parts: a microcontroller, an analog circuit for ECG, and a wireless module. The microcontroller has ap-

proximately 50 % of all power consumption if it does not sleep during an entire exercise or game period. Since the power consumption depends on microcontroller's architecture, operating frequency, voltage, and the amount of computation, it is important to select a microcontroller based on its performance. Therefore, we provide candidates of microcontroller's applicable for the vital signal processing.

Table I Specifications of the microcontrollers

Microcontroller	H8SX1655	PSoC5 CY8C5568	LPC1768
ROM	512KB	256KB	512KB
RAM	40KB	64KB	32KB(local)
A/D channel	8ch	DelSig1ch	8ch
A/D bit	10bit	8 - 20bit	12bit
Current(typ) @3.3V	50mA	37mA @63MHz	53mA
Voltage	3.0-3.6V	2.7-5.5V	2.4-3.6V
Clock on EVB	48MHz	60MHz	96MHz

The microcontroller for the vital sensor node requires at least an I^2C interface for the sensor of tri-axis accelerometer and BT, an A/D conversion channel for ECG, and a UART interface for the wireless module. Furthermore, it requires a sufficient capacity of RAM for the signal processing, and the high speed performance to be able to finish the signal processing within the vital data transmission interval. We list three candidates of 32-bit microcontrollers to satisfy the conditions; the first microcontroller is H8SX1655 of Renesas Electronics, where the development environment is Eclipse + KPIT GNU Toolchain, the second microcontroller is PSoC5 CY8C5568 of Cypress Semiconductor, where the development environment is PSoC Creator, and finally the third microcontroller is LPC1768 of NXP, where the development environment is embed. Table I summarizes the specifications of the microcontrollers. We numerically compared the three microcontrollers in terms of energy efficiency. Note that the source code was written in C or C++ and the software was implemented using compiler and math library of each environment.

Evaluation of Microcontrollers

The evaluation has two steps as follows. In the preparation step, in order to prepare reproducible ECG signal data, some data samples were experimentally recorded to PC files with 10 bits in 100 Hz format. Each sample has 15-minute digital data with 5 sets of 2-minute walk and 1-minute break. In the evaluation step, one of the samples was sent to a microcontroller via the audio I/F of PC. After the reproduced ECG signal is sampled again in 100Hz by the microcontroller, the HR value is calculated every second. In more detail, the ECG signal is passed through an FIR filter, then an HR value is calculated by a time domain threshold method, and finally, an HR value averaged over 10 seconds is output [3].

We implemented the software with both a fixed-point and floating-point computations. The computation time was measured by the internal timer. Here, as the numbers of taps of the FIR filter increases, the more the energy consumption increases, because the microcontroller can go to sleep when it does not have to execute the computation.

We assumed that the sum of the energy consumption at the wireless module and the analog circuit was 930.4 J in 3.3 V 40 mA for 2 hours. Since a 110 mAh battery has 1306.8 J, the rest of the battery (376.4 J) can be used for the microcontroller for 2-hour processing. On the other hand, if small

90 mAh battery is chosen, the microcontroller can use 138.6 J for 2 hours.

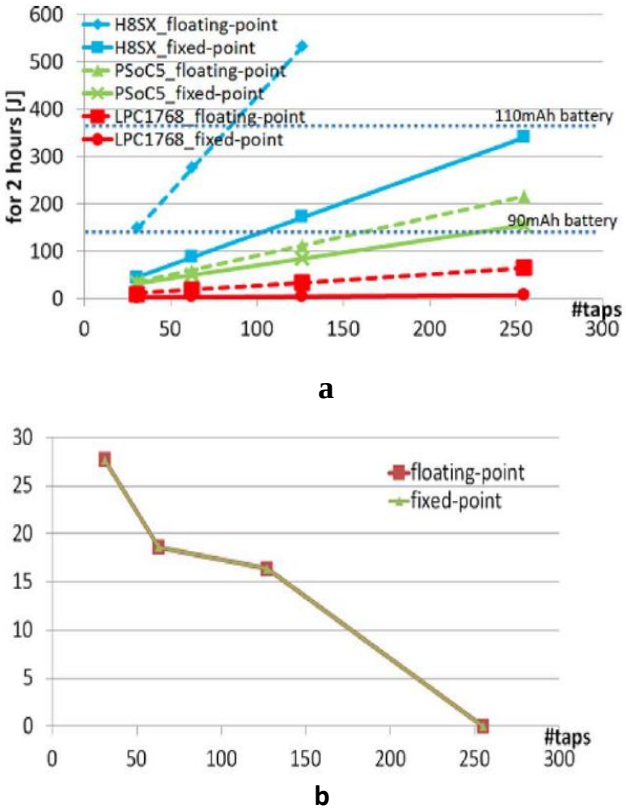


Fig. 2.a Energy consumption for 2 hours versus the number of the taps of the FIR filter, b. The RMSE of HR versus the number of taps of the FIR filter.

Fig. 2.a shows the energy consumption for 2 hours versus the number of taps of the FIR filter. When we implement the FIR filter of 127 taps for floating-point computation, LPC1768 gives the lowest energy consumption of 34 J, PSoC5 gives the second lowest energy consumption of 112 J, and H8SX1655 gives the highest energy consumption of 534 J. The order of microcontrollers is the same for the energy consumption for fixed-point computation, namely, LPC1768, PSoC5 and H8SX1655 give the energy consumption of 5 J, 85 J and 172 J, respectively. Consequently, from the result, we can conclude that LPC1768 is the most energy efficient among the three microcontrollers.

We also studied how the fixed-point computation affected HR. In the HR calculation, rounding off errors are stored in the coefficient parts of the FIR filter, so we compared the root-mean-square error (RMSE) of HR between the fixed-point and floating-point computations. Fig. 2.b shows the RMSE of HR versus the number of the taps of the FIR filter. We can see that the fixedpoint calculation does not affect much the HR calculation accuracy in the region from 31 to 255 taps.

Development of Vital Sensor Node

We have designed and implemented a prototype of the vital sensor node using LPC1768 on a printed circuit board (PCB). Table II shows the specifications of the vital sensor node.

Table ii. The specifications of the vital sensor

Size	60 x 75 mm
Weight	33 g
Wireless frequency	920 MHz
ECG sampling rate	< 100 Hz
Tri-axis acceleration sampling rate	< 100 Hz
Tri-axis acceleration measurement range	± 16 g
Temperature accuracy	±0.5 ^{0} C

Fig. 3 shows the photo of the vital sensor node on the PCB. There are the wireless module in the left, the tri-axis acceleration sensor in the upper left, and the LPC1768 in the center. The upper connectors are for BT sensors, etc.

Fig. 3. Photo of the vital sensor node on the PCB.

Conclusion

In this paper, to realize a wireless vital sensor node workable at a waist position with a small battery, which is applicable for sportsmen in sports exercise and game, we discussed the capability of required microcontroller and evaluated the energy efficiency for three microcontroller candidates. Based on the numerical result, we designed a prototype vital sensor node and finally implemented it on a PCB. Our future investigation is to experimentally evaluate the wireless vital sensor node put to sportsmen during exercise and game.

References

[1] Mucchi, L., Trippi, F., & Carpini, A. (2010, November). Ultra wide band real-time location system for cinematic survey in sports. In Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on (pp. 1-6). IEEE.

[2] GPSports System, "SPIHPU." [Online]. Available: http://gpsports.com.

[3] Masuoka, N., Tsujioka, T., & Hara, S. (2012, October). Heart rate detection from waist ECG/PCG sensors for a vital signal acquisition system for athletes. In Information Theory and its Applications (ISITA), 2012 International Symposium on (pp. 189-193). IEEE..

[4] Chung, W. Y., Lee, Y. D., & Jung, S. J. (2008, August). A wireless sensor network compatible wearable u-healthcare monitoring system using integrated ECG, accelerometer and SpO 2. In Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE(pp. 1529-1532). IEEE..

[5] Gao, T., Greenspan, D., Welsh, M., Juang, R., & Alm, A. (2006, January). Vital signs monitoring and patient tracking over a wireless network. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the (pp. 102-105). IEEE.

[6] Li, C., Wang, Y., & Guo, X. (2010, April). The Application Research of Wireless Sensor Network Based on ZigBee. In Multimedia and Information Technology (MMIT), 2010 Second International Conference on (Vol. 2, pp. 89-92). IEEE..

Available at: www.sabauni.net/ojs

S.J.I.T.N
Saba Journal of
Information Technology
and Networking

Article

N-ATTRIBUTES STOCHASTIC CLASSIFIER COMBINATION FOR ARABIC MORPHOLOGICAL DISAMBIGUATION

Dr.Mohammad Albared,Yemen

Article info

Article history:

Accepted Jan, 2015

Keywords:

Morphological disambiguation,

Classifiers combination,

N-attributes

Abstract

Morphological disambiguation is the ability to computationally determine which morphological tag of a word is activated by its use in a particular context. The main problem in statistical morphological disambiguation of rich morphological languages is data sparseness, where the level of ambiguity is high and the potential tagset size is very large. This paper investigates several fully supervised stochastic morphological disambiguation approaches for morphologically rich languages, with a specific application to Arabic. First, this paper evaluates the direct statistical disambiguation method in which only one tagging model is used. In this approach, each word is assigned a complex morphological tag. In addition, this paper introduces the single-attribute classifiers combination method in which the problem is decomposed into several single-attribute disambiguation subproblems. Then, a classifier combination method, which consists of several trigram HMM tagging models and a module which combines them, is used. Results show that the first method suffers from data sparseness and has large tagging time and the second one has low tagging accuracy. Finally, the paper present a novel approach based on the combination of several N-attributes morpheme-based probabilistic classifiers. First, the morphological disambiguation problem is decoupled into several N-attributes tagging sub-problems. Then, several classifiers are used to solve each subproblem. Finally, the outcomes of all N-attributes classifiers are combined. Several problem decomposition methods and classifiers combination algorithms are investigated. The triple-attributes (N=3) stochastic classifier combination model provides an overall tagging accuracy of 91.5%, reduce the data sparseness problem and saves run time over the direct approaches.

Corresponding author:Dr. Mohammad Albared ,
E-mail address: alialbared@gmail.com

Introduction

Morphological Disambiguation, also known word-class syntactic tagging or fine-grained part of speech (POS) tagging [1-3], is an intermediate layer between morphological and syntactic analysis in which each word appearing in a text is assigned an unambiguous morphological tag [4]. It can also be defined as the ability to computationally determine which morphological tag of a word is activated by its use in a particular context. In morphological disambiguation, the POS tag set is more fine-grained and defined in terms of morphological and grammatical attributes (features) characterizing word structure. Tagging for POS alone would not solve the morphological disambiguation problem[5]. With the term morphological tag, we mean the morphological attributes of a word such as POS, person, number, gender and tense. Tagging with fine-grained tag set is an important step for many NLP tasks such as syntax parsing, word sense disambiguation, semantic parsing analysis and language modeling for speech recognition.

Arabic, like other Semitic languages, has rich inflectional, derivational and templatic morphology [6-8]. The main important qualitative distinction between the typical POS tagging in simple languages and the morphological disambiguation is the large number of possible tags that can be assigned to a word. Unfortunately, the high number of possible tags poses data sparseness challenge for the typical statistical models[1]. A large and fine-grained tag set may enrich the supplied information but the performance of the tagging model may decrease. Morphological ambiguity is very difficult problem in Arabic and some other languages (Levinger, et al., 1995). Thus, finding methods to reduce the morphological ambiguity in the language is a great challenge for NLP researchers . In fact, a much richer model is required to be designed to capture the encoded information when using a fine grained tag set and hence, it is more difficult to learn [9]. In addition, the situation becomes more serious in a low-resource setting.

In general, most previous studies have been treated the morphological disambiguation problem either as a single classification problem and tackled all morphological and grammatical attributes together in one step or as a combination of multiple classification sub-problems, each one is a single morphological attribute classification problem. In the first approach, each word or sub-word (morpheme) in the training data is labeled with a composite tag which can be mapped into a vector of values of grammatical categories"<----------->"[10]. Each morphological and grammatical attribute (morphological properties) has slot "-"of this vector (positional notation). The first slot is usually filled by one of the POS attribute values. Each one of the remaining slot can only contain one of its morphological attribute values or null"-" if its morphological attribute is irrelevant for the POS value in the first slot. The tag set consists of all possible combination of the values of these morphological and grammatical attributes. The potential tag set size is very large especially with

morphology rich languages. However, this approach suffers from the data sparseness. In fact, to ensure statistical significance, a very large training data are required. Even if the training data are very large, it is impossible to avoid the problems of data sparseness and out of word-tag pairs i.e. not all words in the training data have all their possible tags in the training data. Moreover, word-tag or morpheme-tag pairs that appeared with a statistically insignificant frequency will be assigned a poor probability estimate.

In the second approach, the morphological disambiguation problem is decomposed into m disambiguation sub-problems. This is done by dividing the training data into m training data. Although each of these data contains the same words, every one of them is tagged using a different morphological attribute. Then, a classifier combination method, which consists of several simple classifiers and a module which combines them, is used. In fact, this method does not suffer from data sparseness. However, it is less accurate than direct method and it requires a sophisticated combination algorithm to re-impose the linguistic dependencies that is lost during the problem decomposition[11, 12].

In this paper, we evaluate, compare and contrast these two approaches in morphologically rich languages, with a specific application to Arabic. Moreover, this paper proposes a new approach to the morphological disambiguation problem. This method is also based on dividing the morphological disambiguation problem into multiple classification problems. Unlike previous work, the morphological disambiguation problem is divided into several N-attributes classification problems to balance between the sparseness problem and the loss of the linguistic dependencies. Then, each classification problem is solved independently using a trigram HMM classifier. After that, a combination algorithm is used to combine the intermediate results of these classifiers to generate the final result.

The paper is organized as follows: In Section 2, we will review the previous approaches to the morphological disambiguation problem. In Section 3, we present relevant linguistic properties and the morphological ambiguity in Arabic. In Section 4, we introduce the data used and the morphological tag set. In Section 5, we formalize the morphological disambiguation problem in statistical context. In Section 6, we introduce our n-attributes classifier combination approaches and give our experimental results. Finally, we conclude in Section 7.

Related Works

Several works have been developed for morphological tagging of agglutinative or highly inflectional languages such as Turkish, Hebrew and German. In the case of language where the morphology is simple, morphological disambiguation is generally covered under the task of simple POS tagging. The main morphological attributes are embedded in the tag name (for example, Ns and Np for noun singular or plural). In this section, we review related work on morphological tagging for

morphologically rich languages. Turkish words have been actively studied since the seminal work from Oflazer and Kuruoz [13] that used constraint-based approach with hand crafted rules for Turkish morphological disambiguation. They select the right morphological tag based on local neighborhood constraints, heuristics and limited amount of statistics.

Arabic morphological disambiguation has not been studied extensively, especially using statistical approaches. In fact, the main reason is the lack of free publicity morphologically annotated corpora. Recently, a morphologically annotated version of the Quranic Arabic Corpus has become freely available through the Quranic Arabic Corpus project[14]. This facilitates, at least to some extent, the application of advanced machine learning techniques to the problem of Arabic morphological disambiguation. However, Habash and Rambow [15] and Smith et al.[16] worked to address the problem of Arabic morphological disambiguation. These two works utilize Buckwalter Arabic Morphological Analyzer (BAMA). Their approaches are based on disambiguation the output of BAMA i.e. they choose among a limited number of possible tags given by BAMA instead of considering all possible tags that a word may take. However, only Arabic words which can be analyzed by BAMA can be disambiguated. If the correct output is not enumerated by the morphological analyzer, it cannot be predicted.

In this paper, several statistical morphological disambiguation approaches are investigated for morphologically rich languages, with a specific application to Arabic. All these approaches are fully supervised, stochastic and dictionary-free. The proposed methods only uses morphologically-tagged corpus as an information source, and can automatically acquire a knowledge base from this corpus. It can be applied not only to Arabic language but also to other morphologically rich languages. This study first compare and contrast direct classification method and single attribute classifiers combination method. Furthermore, this work also designs a better method to handle the morphological disambiguation which balances between these two approaches. In this method the morphological disambiguation problem is decomposed into several N-attribute classification problems. The main idea of this method is to avoid the data sparseness problem and to retain the linguistic dependencies between the morphological attributes.

Data and the morphological tag set

In this work, the data used is the morphologically annotated version of the Quranic Arabic Corpus[14, 17]. In this version, each word is annotated with its full morphological information including the POS tags of its morphemes. The Quranic Arabic Corpus is an annotated linguistic resource which shows the Arabic grammar, syntax and morphology for each word in the Holy Quran, the religious book of Islam which is written in classical Quranic Arabic (c. 600 CE). The Quranic Arabic Corpus consists of 77,430 words of the Quranic Arabic. In fact, we have used a morpheme-based version which consists of about 128219 tokens. The Qur'an is always published in

fully vocalized versions. For the purpose of this work, we use fully unvocalized versions. All vowels are removed. Nowadays, Arabic is written mostly in unvocalized script. Vowels (diacritics) are no longer used in printed or electronic Arabic text. Unvocalized Arabic texts are more ambiguous than vocalized or partially-vocalized Arabic text. We have chosen the Quranic Arabic Corpus for training and testing our models, since it is the only free available morphologically annotated Arabic corpus. However, our proposed models are language, data and tag set independent i.e. they can also be applied for any language, data and any tag set.

The tag set (morphological tag set), used in this work, consists of multiple dimensions tags. In addition to POS tags, multiple morphological features are assigned to each morphological segment. In the original version, morphological tags are represented in concatenative forms. In this concatenative representation, only the relevant morphological features are shown. Moreover, the default value of some morphological features is not shown such as the definiteness marker "Def". However, as in [10], we represent morphological tags in slots form (-.-.-.-.-.-.-.); a tag consists of 7 slots separated by dots; each position (slot) represents a feature and the tag value (letter, number or three letters) at that location represents a value or attribute of the morphological feature. The dash "-" represents a feature not relevant to a given word. The first slot shows the main POS. The slots 2, 3, 4, 5, 6 and 7 are used to represent other morphological features: person, gender, definiteness, number, tense and voice, respectively.

Problem Definition: The Statistical Morphological Tagging

In this paper, the Arabic Trigram HMM tagger [18, 19] is used as the basic classifier in our environment. In this context, the morphological disambiguation problem is formally defined as follows: given a sequence of words, $w_1^s = w_1, ..., w_s$, find the best sequence of morphological tags $mt_1^s = mt_1, ..., mt_s$ of the words. This can be formulated as follows:

$$mt_1^s = \arg\max_{t_1^s} \prod_{i=1}^{s} p(mt_i | mt_{i-1}, mt_{i-2}).p(w_i | mt_i)$$

MT is the morphological tag set
F_1 is the set of all POS tags
$F_2 = \{1, 2, 3, -\}$
$F_3 = \{m, f, -\}$
$F_4 = \{def, undef, -\}$
$F_5 = \{s, d, p, -\}$
$F_6 = \{perf, impf, impv, -\}$
$F_7 = \{PASS, ACT, -\}$

The simplest way to compute the parameters for the HMM is to use relative frequency estimation, which is to count the frequencies of word/tag and tag/tag. This way is called maximum likelihood estimation. However, data tend to be sparse especially with the large size of the morphological tag set and the small size of the training data. Due to this, the transition probabilities are smoothed using the linear interpolation of unigram, bigram

and trigram maximum likelihood estimates in order to estimate the trigram transition probability.

In any tagging system, we frequently encounter words that do not appear in training data especially when only a limited amount of training material is available. The existence of such words is one of the main problems in any tagging system, since the statistical information of these words is unavailable. Unknown words are usually handled by an exceptional processing. Accuracy of unknown words is usually much lower than that for known words.

For handling unknown words, the Arabic HMM tagger uses a lexical model based on the linear interpolation of both word suffix probability and word prefix probability. Briefly, the model estimates the lexical probabilities of unknown words as follows: Given an unknown word w, the suffix probabilities $P(suffix(w)|mt)$ are estimated using the suffix guessing algorithm as described in (Brant,2000). Then, the lexical probabilities $P(prefix(w)|mt)$ are also estimated using the same way, but, the letters in the words are reversed before adding them to the new word tree in order to find the prefix probability. Finally, we use the linear interpolation of both the lexical probabilities obtained from both the word suffix and prefix to calculate the lexical probability of the word w as in the following Equation:

$$P(w|mt) = \lambda P(suffix(w)|mt) + (1-\lambda)P(prefix(w)|mt)$$

where λ is an interpolation factor.

N-Attribute Classifiers Combination Algorithm

Our methods for morphological disambiguation problem are based on the decomposition of the M N-attributes disambiguation sub-problems by grouping the morphological attributes into M groups. Each group contains N attributes. These methods have to avoid the data sparseness problem and to retain the linguistic dependencies between the morphological attributes.

However, the problem decomposition and m-attribute classifiers combination algorithm is as follow (see Figure 1):

1. First, group the morphological attributes $\{F_1, F_2, F_3, F_4, F_5, F_6, F_7\}$ into M groups $\{G_1, G_2, ..., G_M\}$, where $1 \le M \le 7$ and $G_i \neq G_j$ if i $\neq$ j. Each group contains N morphological attributes.
2. Second, the problem is decomposed into M tagging sub-problems. To do so, M $\{D_1, D_2, ..., D_M\}$ different training data are generated from the training corpus. All training data contains the same words. Each training data D_i is tagged with the possible combination of the morphological attributes of group G_i.
3. Third, M classifiers $\{C_1, C_2, ..., C_M\}$ are used, each classifier is a trigram HMM tagger, and each classifier C_i is trained using D_i.
4. Thirdly, each classifier is tested using a test set that is tagged using the same tag set as in the data used to train it. Each classifier predicts only small set, N, of the seven morphological attributes. Every classifier produces its

own intermediate result which is annotated using the same annotation scheme of the training data used to train this classifier.

5. Finally, the intermediate results are combined using a combination algorithm to produce the final result.

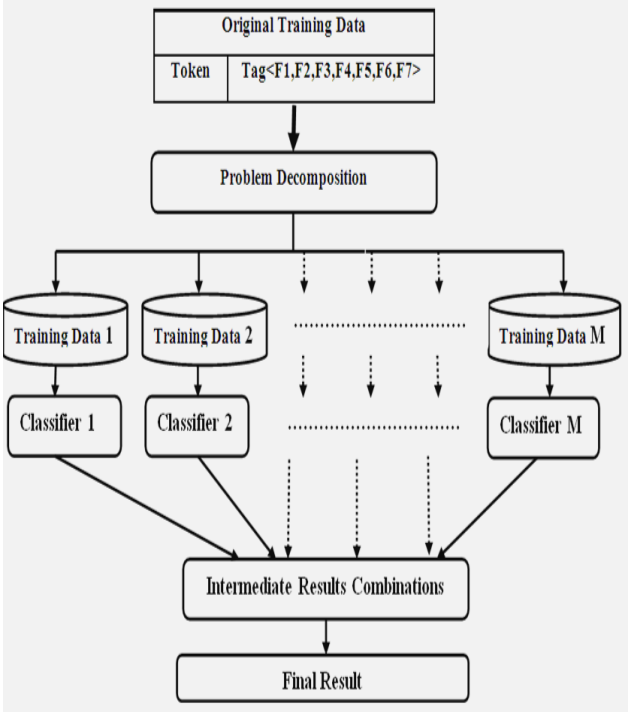


Figure 1 Problem decomposition and classifiers combination algorithm

Evaluation

In the following subsections, we will first evaluate typical HMM POS tagging models for Arabic fine-grained POS tagging (direct classification method) for morphological disambiguation problem. Then, we will introduce several methods of morphological problem decomposition and classifiers combination and evaluate them. In all the experiments described in this paper, the term classifier refers to the Arabic trigram HMM tagging model. In addition, the data sets are split into two sets; training set 89.1% and test set 10.9 %.

Direct Disambiguation Methods

According to Equation (1), standard POS tagging techniques can be used directly to handle the morphological disambiguation problem. However in this experiment, we directly train and test both Arabic HMM tagger with the linear interpolation guessing model and the TnT tagger using the morphologically annotated Arabic Quranic Corpus (M=1). Each word is tagged with a composite tag (N=7). The results of both tagging models are shown in Table 1. The results show that Arabic Trigram HMM outperforms the TnT Tagger. However, the testing time (tagging time) of both taggers is so large due to the tag set size. Essentially the time complexity of trigram HMM-based tagger is $O(n \times T^3)$ [20], where n is the number of words in the target sentence and T is the size of tag set. In this case, the direct method for morphological disambiguation problem T is very large, $T = |MT| \subset \{F_1 \times F_2 \times F_3 \times F_4 \times F_5 \times F_6 \times F_7\}$. The major factor that affects the performance of HMM-based tagger becomes the size of the tag transition probability set. The larger the size of training probability set or tag set, the more time it will take. Therefore it is noticed that when V is the size of vocabulary, and $T = |MT|$ is the number of tags, the space complexity for trigram HMM-based tagger is $O(V \times T^3)$.

Table 1: Results of direct classification methods

Model	Unknown %	Overall %	Training Time	Testing Time
TnT Tagger	33.3	90.33	0.30	23.25
Arabic Trigram HMM with the Linear Interpolation	48.7	91.03	0.30	23.37

Single Morphological Attribute Classifiers Combination

In this experiment, seven (M=7) new training data {D_1, D_2,…, D_7} are created from the original training data. Each one is tagged with the possible values of only single attribute (N=1). The first one is tagged with all POS tags, the second one is tagged with all Person tags {1, 2, 3,-} and so one. Each training data has its own tag set. Then, we use seven trigram HMM classifiers {C_1, C_2,….., C_7}. Each classifier C_i is trained with training data D_i. After that, each classifier is tested using the same test set. Each classifier predicts a small set of possible values of one morphological attribute. The first one is used to predict POS tags; the second one is used to predict person tags and so on. Using the experiment setup in the previous section, the intermediate results of all the simple classifiers are shown in Table 2. As shown in the table, the tagging performance of the trigram HMM classifier (all the classifiers are trigram HMM classifiers.) varies among the morphological attributes. The classifier achieves the highest overall tagging accuracy (98.13%) on voice attribute and the lowest overall tagging accuracy (94.27%) on POS attribute. It is also interesting to note that the unknown word handler algorithm (the linear interpolation guessing algorithm), which has achieved considerable success in guessing the POS of unknown words, has a modest performance in guessing the other morphological attributes of unknown words. In fact, the lexical features used for unknown words POS tags, their suffixes and prefixes, are not appropriate for guessing some of their other morphological attributes. For example, the Arabic nouns are determined or made definite by prefixing the definite article /al-/; by using the noun as first term of an iDaafa (annexation structure); or by suffixing a possessive pronoun to the noun. In Arabic writing, the definite article and possessive pronouns are always attached to the stem. However, because we are using segmented version of the Arabic Quranic Corpus (morpheme-based version), these clitic are separated from the stem and represented as standalone units. In this case, we think that utilizing the lexical features of the context (previous and next words) is more appropriate than the lexical features of the word itself. Moreover, the indefinite markers (suffix sound), which corresponds to the use of “a” or “an” in English, are no longer used in most of the modern Arabic writing (and also we used unvocalized version of the Arabic Quranic Corpus).

Table 2. Intermediate results of all the simple (seven) classifiers

Classifiers and Data		Tag set Type & Size		Unknown %	Overall %	Training Time	Testing Time
C1	D1	POS	45	86.13	94.27	0.0489	0.446
C2	D2	Person	4	82.00	96.27	0.0347	0.0402
C3	D3	Gender	3	75.5	96.11	0.0332	0.0450
C4	D4	Definiteness	3	68.8	95.75	0.0324	0.0430
C5	D5	Number	4	61.6	94.73	0.0336	0.0435
C6	D6	Tense	4	82.24	97.88	0.0316	0.0402
C7	D7	Voice	3	85.4	98.132	0.0309	0.0403

Finally, a hierarchical combinations algorithm is used to combine these intermediate results. The hierarchical combination tries to incorporate more linguistic dependencies which have been lost between the classifiers outcomes. However, not all attributes are relevant to all POS tags. For example, nouns have values for 'definition', 'gender', 'number' and 'person' and can only have 'irrelevant' as the possible value for 'tense' and 'voice''. Conjunctions have no attributes with values other than 'irrelevant'. In this method, we first predict the main POS of the target word, and take this prediction to be true. We then combine the outcomes of only a subset of the other classifiers, determined by the main POS. For example, if the output of the POS classifier is noun, we do not have to look into the output of the classifiers which predict the 'tense' and 'voice'' morphological attributes or even to run them. The results of the hierarchical combination are shown in Table 3. The results, in both Table 1 and Table 3 , show that even we have achieved high single attribute morphological tagging accuracy, the combined results is so low compared to the direct classification. This can be returned knowing that breaking down the morphological disambiguation into many single morphological attribute tagging problems leads to lose a lot of linguistic dependencies which are exist between these attributes, explains this result.

In general, both methods, direct classification method and single-attribute classifiers combination method, have advantage and disadvantages. The direct classification method is more accurate and it retains the linguistic dependencies between the morphological attributes. However, it causes data sparseness and its performance is so slow. On the other hand, the single-attribute classifiers combination method is very fast and it does not suffer from data sparseness problem. The time complexity of single morphological attribute classifiers combination method is small $O(n\times F_1^3)+O(n\times F_2^3)+....+O(n\times F_7^3)+O(n)\approx O(n\times F_1^3)$, where $O(n\times F_i^3)$ is the time complexity of classifier C_i and $O(n)$ is the combination algorithm's time complexity. However, it is less accurate and it does not preserve the linguistic dependencies between the morphological attributes. From that, it is clear that to design a better way to handle morphological disambiguation, we should balance between the two ways. In the next subsections, we describe N-attribute classifier combination methods, where N=2 and N=3..

Table 3: Results of the single-attribute classifiers combination method

Combination Type	Unknown %	Overall %	Testing Time
Simple Combination	30.17	84.67	0.446
Hierarchical Combination	30.66	85.00	0.49

Pair-Attributes Classifiers Combination

The main idea behind this decomposition is that POS attribute is the main morphological attribute and the other morphological attributes are POS-specific. The number of related attributes is fixed for each main POS category. In this method, the original training data is decomposed into six (M=6) new training data {D_1, D_2,…, D_6 }. Each one of these training data is tagged with the possi-

ble values which resulted from the combination of two (N=2) morphological attributes. The six morphological groups are POS.Person, POS.Number, POS.Gender, POS.Definiteness, POS.Tense and POS.Voice. Then, six classifiers {C_1, C_2,….., C_6} are used. Each classifier C_i is trained with training data D_i. Each classifier predicts a small set of possible values resulted from the combination of two morphological attributes (POS and one of the remaining morphological attributes). The intermediate results of all these classifiers are shown in Table 4.

Table 4. Intermediate results of all the pair-attributes classifiers

Classifier and Data		Tag set Type & Size		Unknown %	Overall%	Training Time	Testing Time
C1	D1	POS + Person	52	79.81	93.00	0.052	0.661
C2	D2	POS + Gender	65	67.64	92.83	0.0591	1.224
C3	D3	POS + Definitenes s	49	78.35	93.41	0.051	0.592
C4	D4	POS+ Number	69	63.99	92.69	0.061	1.425
C5	D5	POS + Tense	47	81.02	93.90	0.0513	0.520
C6	D6	POS+ Voice	46	83.70	93.84	0.0500	0.495

Finally, a majority and hierarchical combination algorithm is used to combine these intermediate results. As shown in Table 4, each classifier predicts for each word a POS tag and a value of additional morphological attribute. The predicted POS tags for a word may differ from classifier to another. In this case, we select POS tag which is predicated by the majority. If there is no majority, we select the POS tag that is predicated by the group which includes the classifier whose POS tagging accuracy is the highest. The POS tagging accuracy of each classifier is shown in Table 5. From these results, we have two important observations. First, the POS tagging accuracy, when tagged with additional attribute is better than the POS tagging accuracy when tagged alone (see the result of classifier C_1 in Table 4). This means using additional attributes helps to disambiguate the main POS tag. Second, the majority combination algorithm gives higher POS tagging result than the best classifier. After the POS tag is chosen, we select the output of the relevant attributes' classifiers. For example, if the majority of the classifiers {C_1,C_2,C_3,C_4 and C_5} choose the "ADJ" as the correct POS tags for the target word , we then look into the outcomes of classifiers C_1(for Person),C_3 (for definiteness)and C_4 (for number) and ignore other classifiers. The results of the majority and hierarchical combination are shown in Table 6. The results show that pair-attributes classifiers combination method improves the overall results (4.2%) over single-attribute classifiers combination method. Compared to the direct classification, the direct classification works slightly better than that pair-attributes classifiers combination method. However, the pair-attributes classifiers combination method is much faster than direct classification.

Table 5. POS overall tagging accuracies of pair-attributes classifiers

Classifiers	Overall POS Tagging Accuracy
C_1	94.58
C_2	94.53

C_3	94.28
C_4	94.58
C_5	94.44
C_6	94.33
Majority	95.00

Table 6, Results of the majority and hierarchical combination method (pair-attributes classifiers)

	Tag set Size	Unknown %	Known%	Overall%	Testing time
Majority + Hierarchical Combination	184	42.34	90.82	89.23	2.05

Triple-Attributes Classifiers Combination

In this method, the seven morphological attributes is divided into three groups; each group contains three attributes. They are divided according to the following ideas:

a) POS attribute is the main morphological attribute. Others are POS-specific. So, POS attribute is part of each group.

b) Related or co-occurred attributes are grouped together. So, tense and voice attributes are in the same group.

c) Avoid sparseness i.e. attributes with less possible combinations are preferred to be together. So from the four remaining attributes Person(4 values), Gender(3 values), Definiteness(3 values) and Number(4 values), Person and Gender are selected to be in one group and the other two in another group.

According to that, the seven morphological attributes are divided into three sets, N=3: {POS, Person, Gender},{ POS, Definiteness, Number }and { POS , Tense , Voice }. Then, three, M=3, training data {D_1, D_2, D_3} are created. Each one is tagged with the possible values of one of these sets. We use three classifiers, {C_1,C_2,C_3} where classifier C_i is trained with training data D_i. The intermediate results of the three classifiers on the three training data are shown in Table 7.

Table 7: Intermediate results of all the three-attribute classifiers

Classifier and Data		Tag set Type & size		Unknown	Overall	Training Time	Testing Time
C1	D1	POS + Person + Gender	78	67.4	93.22	0.0672	1.9231
C2	D2	POS +Definiteness + Number	80	60.34	91.92	0.0671	2.0532
C3	D3	POS + tense voice	50	78.59	93.42	0.0506	0.5852

Finally, a combination algorithm is used to combine these intermediate results. In the three-attribute classifiers combination method, each classifier predicts for each word a triple tag (_._._). The first slot is for the POS tag. For the two remaining slots; each one is for a value of one of the two additional attribute. So, we have three POS tags, each one is predicted by one classifier. These three POS tags may be equal or different. If they are different, we choose the POS tag predicted by the classifier (C_1) whose POS tagging accuracy is the highest. C_1 works better than all other classifiers and also works better than the majority voting. So, we do not use the majority voting combination technique. Instead, we directly use

the outcomes of C_1. The POS tagging accuracy of each classifier is shown in Table 8. Moreover, the overall tagging accuracy of each attribute is shown in Table 9. What is important to be noted here is that the POS overall tagging accuracy achieved by C_1 is high (95.20%) compared to most of full statistical models described before, especially C_1, which is trained and tested using data annotated with only POS morphological attributes. This indicates that including some morphological attributes helps clearly to disambiguate the main POS tag. However, after the POS tag is selected; we look of the value of the relevant attributes and ignore others. The results of the combination are shown in Table 10. This result shows that the three (triple)-attribute classifiers combination algorithm achieves comparable accuracy with the direct classification algorithm. In addition, it also outperforms both single-attribute classifiers combination algorithm and pair-attribute classifiers combination algorithm. However, this method is very fast and it does not suffer from data sparseness problem as the single-attributes classifiers combination algorithm and it is accurate as the direct classification algorithm. Moreover, the three-attribute classifiers combination method is much faster than direct classification. The three (triple)-attribute classifiers combination algorithm provides a solution which compromises between the two extremes (the direct classification algorithm and the single-attributes classifiers combination algorithm).

Table 8: Overall Tagging Accuracy of POS Attribute Achieved By the triple-Attributes Classifiers

Classifiers	Overall POS Tagging Accuracy
C_1	95.20
C_2	94.48
C_3	94.44
Majority	95.10

Table 9: Overall tagging accuracy of each attribute achieved by the triple-attribute classifiers

Attribute	The Used Classifier	Overall Accuracy
Person	C_1	97.65
Gender	C_1	97.62
Definiteness	C_2	98.15
Number	C_2	97.00
Tense	C_3	98.72
Voice	C_3	98.78

Table 10 Results of the triple-attribute classifiers combination method

	Tag set Size	Unknown %	Known%	Overall%	Testing Time
Best classifier + Hierarchical Combination	184	60.34	91.79	90.5	2.3

CONCLUSIONS

Morphological disambiguation of Arabic is a hard task due to its rich inflectional, derivational and templatic morphology. The number of potential morphological tags in Arabic is theoretically so large. In this paper, several stochastic methods to statistical morphological disambiguation for rich morphological languages have been presented. These proposed methods do not depend on manually constructed linguistic knowledge such as a dictionary and morphological rules. First, this pa-

per evaluates the direct statistical disambiguation method and the single-attribute classifiers combination method. Then, new methods based on the combination of several N-attribute Trigram HMM classifiers are proposed. In these methods, the multi-attributes morphological tagging problem are decoupled into several N-attributes tagging sub-problems. Then, several classifiers are used to solve each sub-problem. Finally, the outcomes of all n-attributes classifiers are combined. In this way, we try to reduce the data sparseness problem and to retain the linguistic dependencies between the morphological attributes. Among models that we have developed and tested for morphological disambiguation of Arabic, the three (triple)-attribute classifiers combination algorithm achieves a good accuracy, does not suffer from the data sparseness, and saves run time.

In addition, while our experiments are limited to the Arabic language and to specific number of morphological attributes, the models presented thus far in this paper are language independent in nature and applicable to any number of morphological attributes. We believe, therefore, that the experiments will be applicable to other morphologically complex languages.

We believe that these results can be further improved in various ways. First, by improving the performance of the unknown word handling algorithms for each basic classifier by utilizing different lexical features depend on the morphological attributes which the classifier is used to predict their values. The basic classifiers can also benefit from careful tuning of their parameters. Finally, we also believe that linguistic exploration, based on deeper error analysis, and to develop hard constraints which can be used to reduce the error rate of the combination module.

References

[1] D. Hakkani-Tür, K. Oflazer, and G. Tür, "Statistical Morphological Disambiguation for Agglutinative Languages," *Computers and the Humanities*, vol. 36, pp. 381-410, 2002.

[2] H. v. Halteren, W. Daelemans, and J. Zavrel, "Improving accuracy in word class tagging through the combination of machine learning systems," *Comput. Linguist.*, vol. 27, pp. 199-229, 2001.

[3] H. Schmid and F. Laws, "Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-Grained POS Tagging," presented at the Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), Manchester,UK, 2008.

[4] G. Orphanos, "Computational Morphosyntactic Analysis of Modern Greek " Ph.D., Department of Computer Engineering and Informatics, University of Patras, 2000.

[5] M. Levinger, U. Ornan, and A. Itai, "Morphological disambiguation in Hebrew using a priori probabilities," *Computational Linguistics* vol. 21, pp. 383-404, 1995.

[6] N. Y. Habash, *Introduction to Arabic Natural Language Processing*: Morgan & Claypool Publishers, 2010.

[7] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing: Challenges and Solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 8, pp. 1-22, 2009.

[8] M. Y. Tachbelie, "Morphology-Based Language Modeling for Amharic," Ph.D., Department of Informatics, University of Hamburg, 2010.

[9] L. Màrquez, "Part-of-speech Tagging: A Machine Learning Approach based on Decision Trees," Ph.D. PhD Thesis, Departament de Llenguatges i Sistemes Informàtics, Universitat Politčcnica de Catalunya, Universitat Politčcnica de Catalunya, 1999.

[10] J. Hajič and B. Hladká, "Tagging inflective languages: prediction of morphological categories for a rich, structured tagset," presented at the Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1, Montreal, Quebec, Canada, 1998.

[11] D. Shacham, "Morphological Disambiguation of Hebrew," M.Sc., Department of Computer Science, University of Haifa, 2007.

[12] D. Shacham and S. Wintner, "Morphological Disambiguation of Hebrew: A Case Study in Classifier Combination," presented at the Proceedings of EMNLP-CoNLL, Prague, 2007.

[13] K. Oflazer and Ì. Kuruöz, "Tagging and morphological disambiguation of Turkish text," in *Proceedings of the fourth conference on Applied natural language processing*, Stuttgart, Germany, 1994, pp. 144-149.

[14] K. Dukes and N. Habash, "Morphological Annotation of Quranic Arabic," presented at the Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta,19-21 May 2010., 2010.

[15] N. Habash and O. Rambow, "Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop," presented at the Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, Michigan, 2005.

[16] N. A. Smith, D. A. Smith, and R. W. Tromble, "Context-based morphological disambiguation with random fields," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 475-482.

[17] K. Dukes, E. Atwell, and A.-B. M. Sharaf, "Syntactic Annotation Guidelines for the

Quranic Arabic Dependency Treebank," presented at the Language Resources and Evaluation Conference (LREC 2010), Valletta, Malta, 2010.

[18] M. Albared, N. Omar, and M. J. Ab Aziz, "Developing a competitive HMM arabic POS tagger using small training corpora," presented at the Proceedings of the Third international conference on Intelligent information and database systems - Volume Part I, Daegu, Korea, 2011.

[19] M. Albared, N. Omar, and M. J. Ab Aziz, "Improving arabic part-of-speech tagging through morphological analysis," presented at the Proceedings of the Third international conference on Intelligent information and database systems - Volume Part I, Daegu, Korea, 2011.

[20] S. M. Thede, "Parsing and tagging sentences containing lexically ambiguous and unknown tokens," Ph.D., Purdue University, 1999.