HIGHER EDUCATION COMMISSION

H-9, Islamabad (Pakistan)

Technology Development Fund

TECHNOLOGY DEVELOPMENT FUND FOR Ph.D. SCHOLARS
PROTOTYPE/PRODUCT DEVELOPMENT GRANT PROPOSAL
COVER SHEET FOR PROPOSAL

For HEC use only
Proposal Identification Number

A1. TITLE OF PROPOSED PROJECT

FDCBA: Flaw Detection and Correction for Bio-informatics API's

A2. CLASSIFICATION/NAME OF PRODUCT/PROCESS/SERVICE TO BE CREATED

Automated detection and correction of Antipatterns in existing REST Web Services.

A3. List of Content:

B1. RESEARCH DOMAIN

☑ Product Development/ Improvement ☐ Process Development/ Improvement

B2. STATE FIELD OF RESEARCH AND SPECIALIZATION (For example; PREVIOUS EXPERIENCE WITH INDUSTRIAL RESEARCH)

Major/Specialization Information Retrieval Relevant Industry Health and Medication

C1. PROJECT DIGEST. Describe the proposed PROJECT using (about 300) in easy business language.

The purpose of this project is to develop a tool suite that dynamically detect and correct the flaws in Bioinformatics APIs that helps to improve the accuracy of diseases like Blood cancer, bone marrow and irregular protein values in human blood. This application will help the hospitals working in the area of Bioinformatics to improve the accuracy of their systems after removing flaws in already available systems. This will help doctors, practitioners and specialist for improvement in diagnosis for cancer patients after providing online platforms. This platform will bridge the gap between existing system used in Pakistan as well as practitioners working in United States of America, Canada, and Japan etc. This application is free and can be used by any medical research lab for improved diagnosis after providing communication channels by linking researchers from different countries. The application will have different modules that helps clinical staff in decision making. The modules first detect flaws in existing systems already available and then correct them dynamically with a comparison of accurate report generated by the system before and after the flaws detection and correction for bio informatics APIs. This will not only assist practitioners but also researchers working in this domain after providing generalized information both for medical community as well as academia. The application also uses Natural Language Processing(NLP) technique that points the flaws in already available systems and translate them in natural language easily understandable by common users.

C2. Patent/IP Search:

1. What is status of your idea vis-à-vis similar patents?
   No Patent

2. Have you already filed a patent application?
   No

3. If a patent already exists on your idea/concept, then what is your strategy and rationale in making the proposed research?
   NA

D. PRINCIPAL INVESTIGATOR (from University / Institution)

D1. PRINCIPAL INVESTIGATOR NAME (full with no initials)	D2. HIGHEST DEGREE	D3. POSITION
Fatima Sabir	PhD	Associate Professor

D4. DEPARTMENT/SECTION	D5. UNIVERSITY/INSTITUTION	D6. CNIC#3620377147699
Al-Khwarizmi Institute of Computer Science	University of Engineering and Technology, Lahore	Please attach a crossed copy 

D7. Telephone:(area code, number and extension)
Mobile: +92 313 6192000 Fax: (area code, number) :
Email: usman.ghani@kics.edu.pk

Postal Address:
Associate Professor
Al-Khwarizmi Institute of Computer Science
University of Engineering and Technology, Lahore

E: INDUSTRIAL PARTNER (from Collaborating Industry)

E1. Industrial Partner NAME (full with no initials)	E2. HIGHEST DEGREE	E3. POSITION
Umair Abdul Hai	B.Tech Honor	Assistant Director
E4. SECTION / UNIT	**E5. FACTORY / INDUSTRY**	**E6. OFFICIAL MAILING ADDRESS**
Purchasing		Office # 17-G, Al-Hafeez Shopping Mall Mian Boulevard Gulberg III, Lahore ,Pakistan

E7.Partner Industry Certification/Registration:

NTN/STN: 2802017-7

E8. Telephone: 042-35774853 Fax: 042-35771794
Mobile: 0331-4477102 Email:Chishti_sons@yahoo.com

F1. PROPOSED DURATION OF PROJECT:(*inmonths*)	F2. PROPOSED STARTING DATE
24 Months	1st July 2018

F3. TOTAL FUNDS REQUESTED	F3.HEC COMPONENT (UPTO 14Million)	F4- INUDSTRY COMPONENT
RS. 6.699 million	A. 1st YEAR installment RS. 4.304(Million) B. 2nd YEAR installmentRS.2.395 (Million)	A. 1st YEAR installment RS. 0 (Million) B. 2nd YEAR installment RS. 0 (Million)

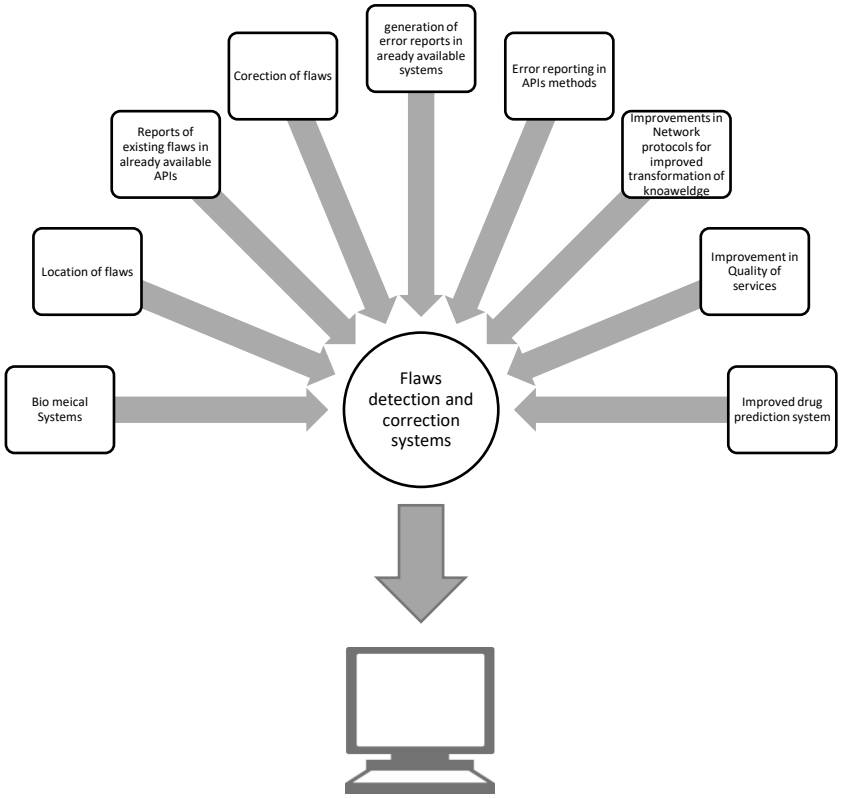
SIGNATURE OF PRINCIPAL INVESTIGATOR	SIGNATURE OF PARTNER INDUSTRY
Date 13/11/17	Date

ENDORSEMENT OF THE HEAD OF INSTITUTION (Vice-chancellor/Rector of University, Director of Degree-awarding Institutions)	ENDORSEMENT OF THE HEAD OF INDUSTRIAL ORGANIZATION
Signature & Date Name & Title: Prof Dr. FAZAL AHMED LODHI VICE CHANCELLOR University of Engineering and Technology, Lahore Address: University of Engineering and Technology, Lahore Phone: 04299029201 E-mail: vc@uet.edu.pk Fax: 04299029202 Date:	(Must be stamped) Signature & Date: 13-11-2017 HAFSA INTERNATIONAL LAHORE Name & Title: M. Ahmad Address: Office #LG-17, Al-Hafeez Shopping Mall, Main Boulevard Gulberg

PROJECT DETAILS

1. PROJECT SUMMARY

The primary goal of this project is to improve the Biomedical APIs that assist practitioners and researchers for improved diagnosis after detection of flaws using Natural language processing techniques with the help of Artificial intelligence. First step is the detection of flaws in Biomedical APIs and then dynamic correction of these flaws that helps to improve the accuracy of disease after improving flaws in these web APIs. It is common practice to send clinical reports for examination across the borders that help local practitioners for improvement in diagnosis. There are many web services used for the communication in Biotechnology but still there are many techniques reported at software side for the reporting and diagnosis for cancer patients, general medical prescriptions and DNA analysis. It is a global practice that people search about diseases over that internet and get information or even prescription knowledge after greeting information over the web. Therefore, the major service providers in the field of bio medical sciences have flaws that may affect the system too and even reduce the accuracy of diagnosis. So this tool suite can be used by industry for the improvement in their already available systems as well as correction in their systems after pointing flaws and dynamically correct the systems that generate more accurate reports as compare to the previously available reports. These reports are generated in comma separated values that points the flaws as well as points the location of flaws detection as well as corrections



The application will help to point the services defects that stop the users for getting in time information for various issues. This application will not only point services errors but also correct these issues and helps common users, para medical staff for the between communication without any delay. This application also uses NLP (Natural Language Processing techniques) to improve and understand better readability of web services. We also help common users to not only understand the address of specific web services used in Bio informatics but also helps major service providers to remove the issues in their services. Overall, the purpose of this application is to design and implement a detection and correction approach for clinical support system for all types of APIs used for mobile applications as well as web.

2. PROPOSED GOALS/OBJECTIVES (please identify quantifiable goals)

The goal is to develop a tool that detect the problems in available application from the field of BIO informatics and especially used in mobile applications that stop common users for using these services as services are difficult to access due to some services errors like 420(service not available) or 404(Service time out errors). There are also some issues as users wants to download report but service providers not give rights for this. Users want to send report in specific format but this facility is also not available that stop practitioners and a user to get benefit. Therefore, our system aims to solve the service provider's errors that improve the accuracy of the web services and provide better facility too. The services will be more easy to use and can be utilized well. The system will not only help practitioners, lab researchers but also common users to get better facility. The end user of this product can be any doctor/ lab supervisors directly involve in report and common users who want to send and retrieve information too.

GOALS/OBJECTIVES

1. To provide better facility to service providers working in the field of Bio-Informatics
2. To remove the service protocols errors that stop users for using services.
3. To assist clinicians for improvements in diagnosis.
4. To bridge the gap between industry and academia after providing improve retrieval of services.
5. To save time for doctors and practitioners
6. To facilitate medical industry for between service availability.
7. To provide timely access of all type of web services for examine patients.
8. To help common users for better understanding of their health issues.
9. To improve the economy of Pakistan as service providers can remove their services errors.
10. To stop any clinical errors that occur due to errors available in APIs.

3. INTRODUCTION (not to exceed one page**)**

SBSs (Service Based systems) are continuously evolving to meet the rapidly changing user requirements and technology variations for improved performance. These changes affect the evolution of SBSs that may cause the degradation in design and quality of services, commonly called as Antipatterns. SOA (Service Oriented Architecture) is widely used in software industry due to its scalability, flexibility and usage of platform independent, autonomous services that can be used via internet [1].

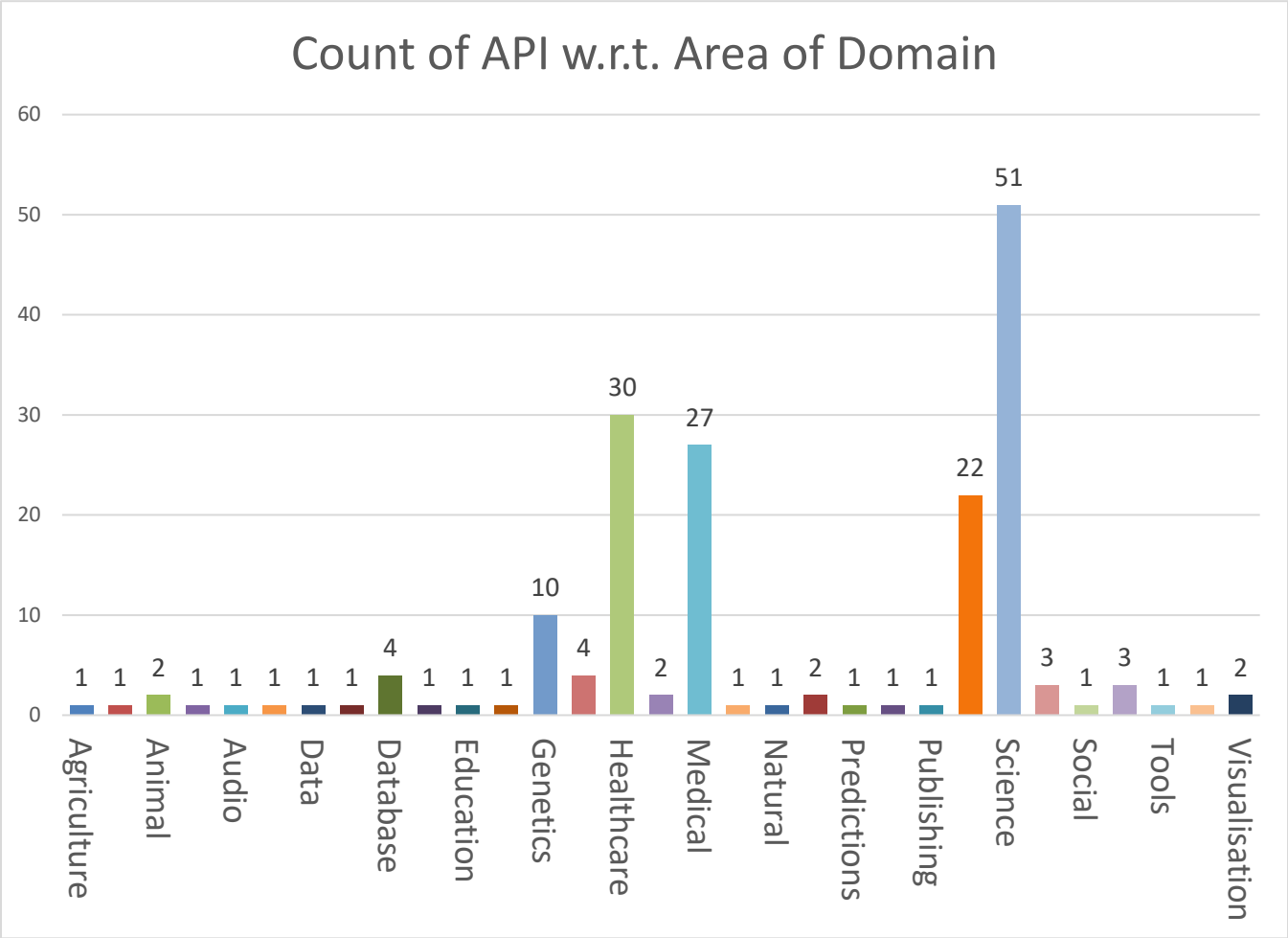
Service based systems can be implemented using wide range of SOA technologies, such as SOAP, REST, RPC, OSGi, Web services, Java RMI, WCF and Apache thrift. SOA allow us to build different types of systems like Google map, Amazon, EBay, PayPal, FedEx etc. Service based systems are also called composite systems, which are composed of autonomous web services that are controlled by the different composition processes [2].

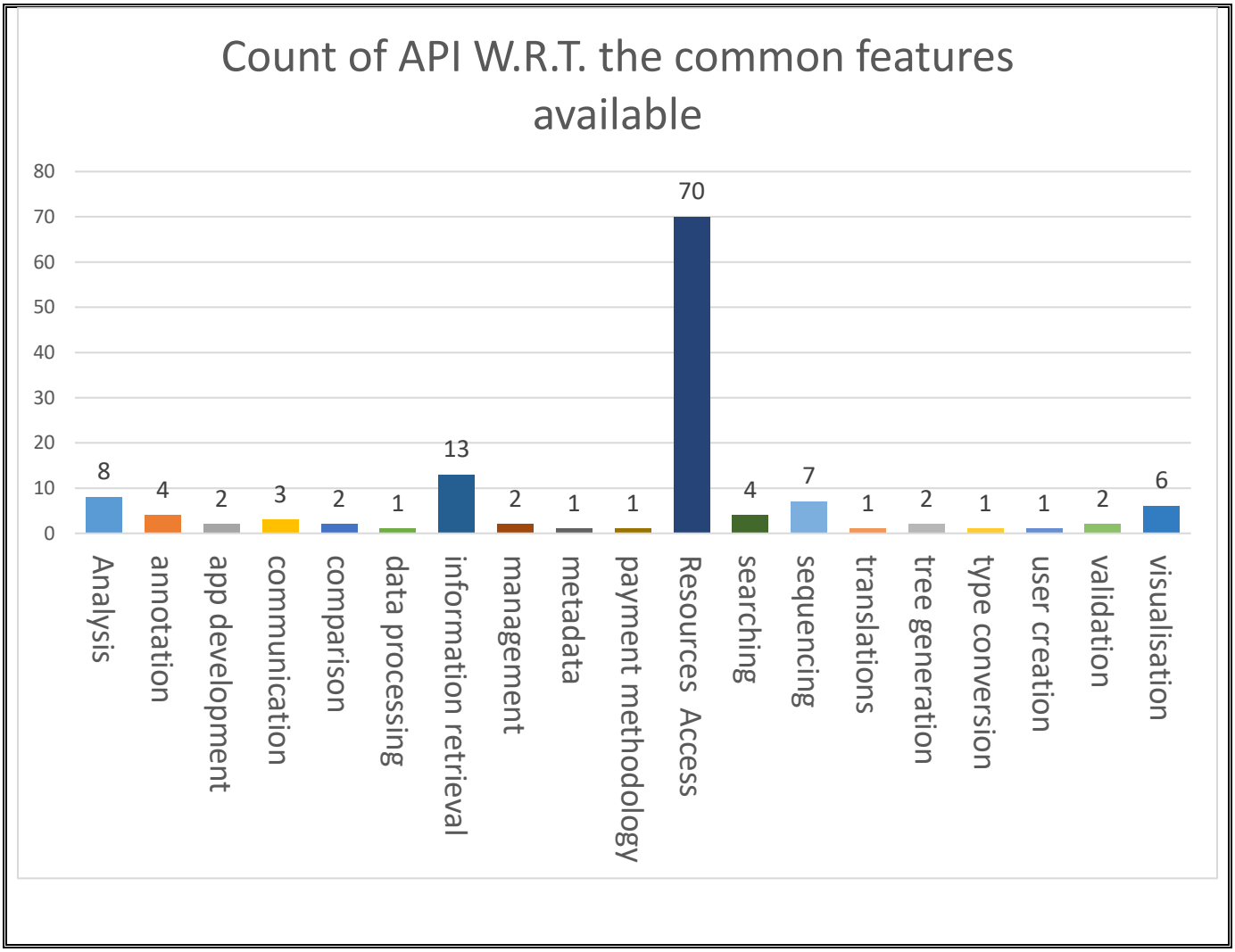
Due to the rapidly growing SBSs industry, the underlying proposed study aims to detect antipatterns/flaws in different SOA technologies especially used for Bio Informatics[3][4] that deteriorate the performance of Web services. As catalog of antipatterns/flaws for SOA technologies is not large, only few antipatterns are reported for SCA, REST and SOAP. Moreover, there is no work reported for the correction of REST services in the field of Bio Informatics. Our focus is to provide detection algorithms for different SOA technologies particularly for the web services used in Bio Informatics, REST full web services and SCA by improving accuracy of detection algorithms. We also intend to propose correction algorithm for REST services that is still not reported in literature and helps medical practitioners for better service retrieval and availability.

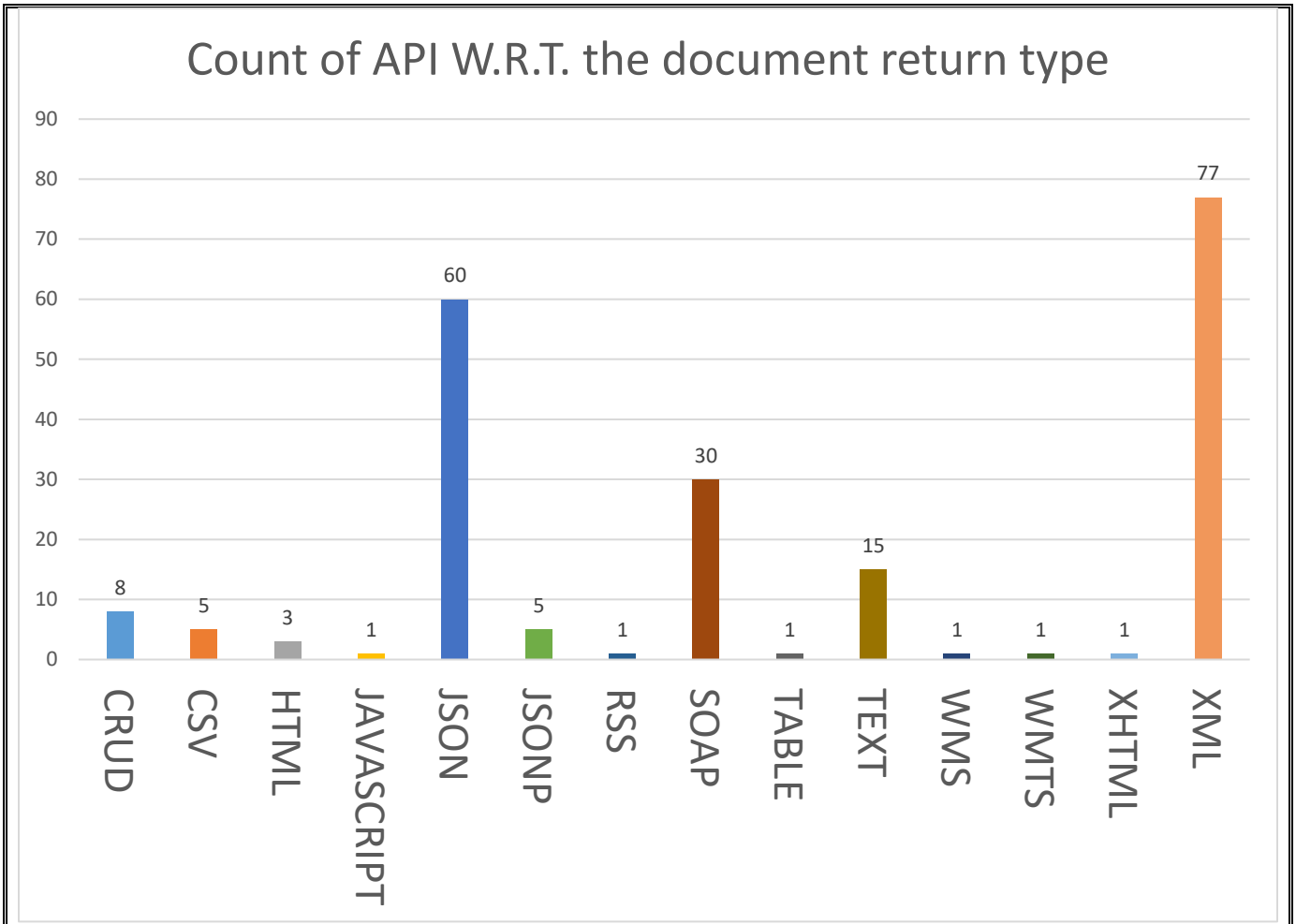
4A. BACKGROUND AND METHODOLOGY OF THE PROPOSED RESEARCH (Not to exceed two pages**)**

The focus of our research is to detect different flaws in SCA, Web services, and REST full Web services. We will also analyze the effect of Antipatterns (flaws) on overall system performance that will help software developers to optimize

the development cost by reducing the time for maintenance as well as efforts. The detection of Antipatterns from SBSs also helps to improve the services. However, flaws detection systems are only available for services working for social apps like Facebook, YouTube and Twitter but there is no such work reported for the field of bioinformatics Moreover, the proposed research intends to detect flaws from the services and from different open source, SBSs used in medical field.




Count of API W.R.T. the common features available
80
70
60
50
40
30
20
10
0
8
4
2
3
2
1
13
2
1
1
70
4
7
1
2
1
1
2
6
Analysis
annotation
app development
communication
comparison
data processing
information retrieval
management
metadata
payment methodology
Resources Access
searching
sequencing
translations
tree generation
type conversion
user creation
validation
visualisation



4B. RESEARCH PLAN: SCHEDULE/PHASING (Preferably with a time-chart not to exceed one page)

Phase	***Duration***
Requirement Analysis	**3 months**
Application Design	**3 months**
Architecture Design	**3 months**
Modules Implementation Phase for Web services	**4 months**
Integration of Modules	**4 months**
Testing and Debugging	**3 months**
Deployment	**1 month**

4C. REFERENCES (cited in 3, 4A & 4B; not to exceed two pages)

1. Erl, T. (2005). *Service-oriented architecture: concepts, technology, and design*. Pearson Education India.

2. Spanoudakis, G., & Mahbub, K. (2004, September). Requirements monitoring for service-based systems: Towards a framework based on event calculus. In *Automated Software Engineering, 2004. Proceedings. 19th International Conference on* (pp. 379-384). IEEE.

3. Siegel, Rebecca L., Kimberly D. Miller, Stacey A. Fedewa, Dennis J. Ahnen, Reinier GS Meester, Afsaneh Barzi, and Ahmedin Jemal. "Colorectal cancer statistics, 2017." *CA: a cancer journal for clinicians* 67, no. 3 (2017): 177-193.

4. Chojnacki, Szymon, Andrew Cowley, Joon Lee, Anna Foix, and Rodrigo Lopez. "Programmatic access to bioinformatics tools from EMBL-EBI update: 2017." *Nucleic acids research*45, no. W1 (2017): W550-W553.

5. IMPACT

The project will effect mainly the health sector of Pakistan by improving service availability for health sector. It is now global practice to use major services as medical students as well as doctors also get benefits from the service providers. Therefore, there is need for improvement in service availability as well as efficiency for improved diagnosis. The project will make use of local intellectual resources for the improvement in APIs

1. The tool will be publically available and can be used by industry to check the errors in their systems.
2. The project will reduce the cost after providing timely data to end users.
3. It will improve the health of Pakistan by utilizing Information Technology

6. Sustainable Development Goals (SDG's) (How and which of the SDG's will be addressed in this study? Justify how the proposed research will contribute to achieve SDG's of Pakistan. For details on SDG's /s please visit:)

- https://www.programmableweb.com/news/72-medical-apis-avvo-national-library-medicine-and-nhs/2012/05/16
- **https://datascience.nih.gov/bioCADDIE**
- **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4941838/**
- https://www.pharmamanufacturing.com/articles/2009/065/

- **Goal1:** The project will ensure improved diagnosis system for cancer patients by removing logical errors in their APIs that stop doctors to consult in different areas.
- **Goal 2** This project will assist companies working in the field of Biomedical for improve information availability and usability.
- **Goal 3:** The project can be used locally by supporting open source systems used in Biomedical sciences working in distributed environments.
- **Goal4:** The project plays a significant role by facilitating researchers and scientist working in the field of cancer research for improved service availability by sharing their resources and getting timely feedback for improved diagnosis.

7. PROJECT PARTNERS (information on Industry)

Please give a brief introduction of the collaborating industry, especially information on turnover, import/export profile, stock exchange listing etc. Please indicate the portion of the proposed research program to be carried out at the Partners organization. Also state that how and where the Partner's budgetary contribution will be utilized.

8. PROJECT Business Plan/Work plan (Attach the Business Plan for your proposed project)

Please find attached business plan

9. PROJECT OUTPUT

The project aims to build tool used by web service provides working in the field of Bio-informatics. The main feature is the dynamic analysis of the application that helps to improve the effective service availability. There are two major market products of our project: 1) Any type of bio-informatics service providers for their better service availability can use prototype tool. 2) Results are available in the form of excel sheets that can be used to mine the data for particular type of errors associated with Bio-Informatics web services.

10. FACILITIES AND FUNDING

10A. Facilities: equipment available for the research project IN THE HOST UNIVERSITY/INSTITUTION & THE COLLABORATING OGRANIZATION
10B. Scientific Personnel (at the PI institution) a. Available b. Required* *Involvement of research students is encouraged.
10C. Other funding available for the proposed studies (if any)

11A. PRINCIPAL INVESTIGATOR

A brief resume of research accomplished in the last 05 years. Please specify title of the research proposal(s), duration, funding source(s) and award amount(s). Detailed CV can be placed as annex at the end.	
1. Please attach C.V.	
2. Number of Publications during the last five years & page numbers on the C.V. where these publications are listed	National: _____ Please see pages_____ of CV International: _____ Please see pages : _____ of CV
3. Number of research projects completed & page number where this information appears	Basic: _____ Please see pages _____ of CV Applied: _____ Please see pages_____ of CV

11B. Industrial Partner (Profile of Partner industry, Accreditation and Certification, Website, Focal Person Contacts. Email, mobile and landline**)**

CV is attached with the proposal

11C. CO-PRINCIPAL INVESTIGATOR

CV is attached with the proposal

12A. ESTIMATED BUDGET FOR THE PROPOSED RESEARCH PERIOD (Rs. in million, please avoid simple calculations)

HEC will fund up to Rs. 14 Million and there is no limit for Industry but proposals with industrial financial assistance will be preferred. Please submit M.S. Excel sheet separately.

DESCRIPTION	YEAR 1		YEAR 2		Total Amount	
	HEC	Industry	HEC	Industry	HEC	Industry
A. Salaries and Honorarium						
PI: One month/year of basic salary @	**0.085**	**0**	**0.085**	**0**	**0.17**	**0**
CO-PI: One month basic salary for whole project @	**0.085**	**0**	**0.085**	**0**	**0.17**	**0**
Rs=25000/- per month fixed for Ph.D. (x students)	**0.9**	**0**	**0.99**	**0**	**1.89**	**0**
Rs=15000/- per month for M.S./M.Phil. (x students)	**1.08**	**0**	**1.188**	**0**	**2.268**	**0**
Subtotal A:	**2.15**	**0**	**2.348**	**0**	**4.498**	**0**

Prototype Development Cost

B. Permanent Equipment						
Item Detail & Quantity	**HEC**	**Industry**	**HEC**	**Industry**	**HEC**	**Industry**
Laptops x 2	0.2	0	0	0	0.2	0
Desktop Computers x 9	0.9	0	0	0	0.9	0
Printer	0.05	0	0	0	0.05	0
GPU x 6	0.6	0	0	0	0.6	0
Data Processing Servers x 3	0.3	0	0	0	0.3	0
Subtotal B:	**2.05**	**0**	**0**	**0**	**2.05**	**0**
C. Expandable Supplies						
Subtotal D2:	0	0	0	0	0	0
D3. Miscellaneous	**HEC**	**Industry**	**HEC**	**Industry**	**HEC**	**Industry**
Audit Fee (Max. Rs 10,000)	0.01	0	0	0	0.01	0
Accountant Fee (Max Rs. 10,000)	0.01	0	0	0	0.01	0
Subtotal D3:	0.02	0	0	0	0.02	0
Subtotal (D1 + D2 + D3):	0.02	0	0	0	0.02	0
E. Indirect cost (University overheads)						
15% to University ORIC or 02% of Total direct costs to the university portion meet office support, utilities, etc.)	0.0844		0.04696		0.13136	
Grand Total (A + B + C + D + E):	4.3044	0	2.39496	0	6.69936	0
Total Budget HEC + Industry Components	4.3044		2.39496		6.69936	
Total Allocation to Research Project By HEC	**4.304**		**2.395**		**6.699**	
Total Allocation to ORIC (If any) By HEC						
Grand Total	**4.304**		**2.395**		**6.699**	

13. JUSTIFICATION (Please justify your request in a background of the existing facilities available at the host Institute.)

A. **Salaries & Allowances** (All positions, other than PI and industrial partner, must be fully justified. Please give qualifications/requirements of each of the new full-time positions requested for in the Proposal.)

Principle Investigator: Overall monitoring and execution of the project. Responsible for reporting progress of the project to HEC.

Co-Principle Investigator (Co-PI) Providing expertise in natural language processing and smooth functioning of the project.

PhD students:

1) Detailed monitoring of the overall project. Work on natural language Processing part of the project.
2) Work on semantic web part of the project. Will develop APIs for SW integration.

MS Students:

i. Develop software, algorithm and testing of each and every module in natural language processing
ii. Will work on semantic web for generating ontologies
iii. Integration and APIs development

B. **Permanent Equipment** (Please identify major items (over Rs. 25,000). Major pieces of equipment costing over Rs. 0.1 million must be fully justified. Minor items (under Rs. 25,000) may be lumped into one.)

The laptops and desktop computers are required for development. Printer is required for documentation. GPU are required for deep learning. One server will be hosting the web application. The second server will be for database and third will be the replication of second server.

C. **Expendable supplies**

D. **Other Costs.** (Travel must be justified.)

14.

(A) Enlist Three **Academic Evaluators/Experts** from Higher Education Institutions/R&D organizations/Universities (Name, Mobile number, postal and email address information)

1. Dr. Abdul Rehman, (0333-8418450)

Dept. of Computer Science, University of Gujrat (Hafiz Hayat Campus), Gujrat

a.rehman@kics.edu.pk

2. Dr. Junaid Arshad (0333-4400478)

Dept. of Computer Science and Engineering, UET, Lahore

junaidarshad@uet.edu.pk

3. Dr. Muhammad Afzal (0300-4245770)

Dept. of Computer Science and Engineering, UET, Lahore

shmafzal@yahoo.com

(B) Enlist Two **Industrial Evaluators/Experts** from Industry/Government and Private R&D organizations (Name, Mobile number, postal and email address information)

1. Imran Sarwar (0321-4456790)
General Manager, SPARCO, Islamabad
Imran.sarwar@kics.edu.pk

2. Mamoon ur Rasheed (0322-4349020)
Senior Analyst,
1-C Block N, Model Town Extension=Lahore 54000
mamoonurrasheed@gmail.com

(B) Enlist Two **Industrial Evaluators/Experts** from Industry/Government and Private R&D organizations (Name, Mobile number, postal and email address information)

1. Imran Sarwar (0321-4456790)
General Manager, SPARCO, Islamabad
Imran.sarwar@kics.edu.pk

2. Mamoon ur Rasheed (0322-4349020)
Senior Analyst,
1-C Block N, Model Town Extension, Lahore 54000
mamoonurrasheed@gmail.com

Signature Principal Applicant

Date: 13/4/17

Checklist:

Make sure the following must accompany the application otherwise proposal will not be shortlisted.

1. 3 hard copies with proper Tape binding (no spiral binding will be accepted) ✓ Y/N
2. Soft copy must be emailed to gsarwar@hec.gov.pk. ✓ Y/N
3. Both hard and soft copy must reach well before the deadline. ✓ Y/N
4. Budget submitted on prescribed M.S excel sheet sent with soft copy. ✓ Y/N
5. TDF-Proposal information sheet on prescribed M.S excel sheet sent with soft copy ✓ Y/N
6. Application must be routed through University ORIC or Directorate of Research ✓ Y/N
7. CNIC copy of Principle Investigator ✓ Y/N
8. Passport size photograph ✓ Y/N
9. Industrial Support letter from partner industry ✓ Y/N
10. Partner Industry NTN/STN provided ✓ Y/N

11. Industrial Partner certification (Registration, ISO or any other etc.) ✓ Y/N
12. Mobile/cell phone number and personnel & official email ✓ Y/N
13. Authentication from Head of the institution and submission through ORIC ✓ Y/N
14. Authentication from Head of the industry not necessary on the form (a separate letter on industrial letter pad will serve the purpose properly addressed to head of university/institution) ✓ Y/N
15. CV of PI ✓ Y/N
16. CV of Co PI ✓ Y/N
17. CV of Industrial Partner ✓ Y/N
18. All documents (required) must be in one application file (soft and hard). ✓ Y/N
19. Quotations of the equipment & Supplies (if cost is more than PKRs.0.1 million) ✓ Y/N
20. Project Key Performance indicators against which performance will be evaluated ✓ Y/N

HAFSA INTERNATIONAL

Sales Office:
Office # 22, 2nd Mezzanine Floor,
Al-Hafeez Shopping Mall,
Main Boulevard, Gullberg III, Lahore.
Ph: 042-35771795-6 Fax: 042-35771794
E-mail: shakeel-duplo@hotmail.com
http://www.hafsainternational.com

Head Office:
852/2 Gulistan Colony Lahore. Ph: 0300-4186839

GST#: 0310900900237 NTN#: 3010591-9

To Whom It May Concern

Subject: Research Collaboration with Al-Khwarizmi Institute of Computer Science, UET Lahore

The University of Engineering and Technology, Lahore encompasses Al-Khawarizmi Institute of Computer Science (KICS), which has been playing an important role in research community for years. The said research institute has recently proposed a prototype system, "**PMAd: Portable Medical Advisor**" which will contain following major features:

- Clinical Decision Support System helping doctors in their decision making process
- Provision of Dosage Calculators, IV drip rates calculators, lab test details etc. to reduce medical errors
- Differential Diagnosis
- Illustrative procedural guidelines and symptoms for better diagnosis
- Provision of algorithms for basic life-saving protocols like BLS, ACLS etc.

We found "**PMAd: Portable Medical Advisor**" very encouraging and suitable for Pakistani community as previously no such initiatives have been taken. Based on effectiveness and need of such system in the health infrastructure of Pakistan, it is recommended that this project should be supported. We are interested in the provision of professional and technical assistance, since successful implementation of this project will help saving many lives.

We hope this research work will bring the technology of clinical decision support system in Pakistan and help doctors and medical practitioners in saving many lives.

HAFSA INTERNATIONAL LAHORE