



Nafees Natalique – Character Based Nastalique Font for Urdu
Grant No. 0201A6_L011

Final Narrative Report

Character Based Nafees Nastalique Font for Urdu

October 3, 2003

**CENTER FOR RESEARCH IN URDU LANGUAGE PROCESSING
NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES, LAHORE
PAKISTAN**

Acknowledgements

This work has been funded by ICT R&D Grants Programme: Asia Pacific (Ref 0201A6_L011), administered through Asian Media Information and Communication Center (AMIC) and funded through International Development Research Center (IDRC) of Canada, Asia-Pacific Development Information Program (APDIP) of UNDP and APNIC.

Contact

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences
852 B Faisal Town, Lahore,
Pakistan.

Phone: +92 42 – 111 128 128

Fax: +92 42 – 516 5232

Email: crulp@nu.edu.pk

URL: www.crulp.org

Synthesis

Traditionally, Urdu has been written in Nasta'leeq script – the most frequently used writing style for Urdu. This script is cursive having a complex and context-sensitive structure. This project has aimed to perform a quantitative analysis of Nasta'leeq rules. Nasta'leeq itself is written in various styles dictated by prominent calligraphers.

This font is developed according to calligraphic rules, following the style of Syed Nafees Al-Hussaini (Nafees Raqam) (www.calligraphyislamic.com/Profiles/Nafees.html), who is one of the finest calligraphers of Pakistan. Guidance and calligraphy of basic glyphs for the font has been provided by Syed Jameel-ur-Rehman.

The project utilizes Open Type Font (OTF) specification to enable the realization of context-sensitive writing systems like Urdu Nasta'leeq. This project presents a scientific study of Nasta'leeq orthography and models it using OTF specification. Output of this project is the model of these rules in the form of character based Nasta'leeq font for Urdu. This font captures context sensitive substitution grammar of Nasta'leeq, formulates its cursive positioning rules and derives optimized *nuqta* (marks on alphabets) placement rules.

This project has been developed by following steps of conventionally used software engineering life cycle, including:

1. Domain Understanding and literature review
2. Requirements gathering (output in the form of documentation of logical model)
3. Prototype development (output in the form of Nafees Naskh OTF font)
4. Design phase (output in the form of documentation of physical model)
5. Implementation
6. Testing
7. Release

On successful completion, it has become easy to disseminate information in Urdu language through electronic media including internet. Development of this font has enabled the users to publish electronically in Urdu and thus reach out to the extensive readership across the world. Furthermore, through various documents as specified above, this project has quantified Nasta'leeq rules to significant detail and analyzed methods for modeling and rendering complex fonts.

Research Problem

The major research problem addressed in this project was modeling of character based Nasta'leeq font for Urdu. This requires quantitative analysis of Nasta'leeq rules including proper contextual substitution of shapes, their cursive positioning and correct *nuqta* placement.

Research Finding

In quest to model Nasta'leeq we sought to learn its underlying principles including calligrapher's intuition. Through series of lectures on Nasta'leeq, long discussions with calligrapher and analysis of hand written books we matured our perception of Nasta'leeq's inherent architecture. Thus, were able to enhance our 'knowledge base' of this writing style. At the end of this practice we succeeded in consolidating Nasta'leeq and producing its logical model as we understood it. Major research findings include:

1. Capturing context sensitive substitution grammar of Nasta'leeq
2. Formulating its cursive positioning rules
3. Deriving optimized Nuqta placement rules, and
4. Implementing kerning rules where necessary

The above findings have also been documented in logical and physical model of Nafees Nasta'leeq.

Towards the end of our effort, we faced major limitations in OTF specification (size of GPOS table of OTF specification was exhausted). Although we opted for optimal solutions but through this exercise we eventually tested existing technology for font development, identified its limitations (with reference to Nasta'leeq font) and indicated areas where enhancements are needed.

Fulfillment of Objectives

This project utilizes Open Type Font (OTF) specification to enable the realization of context-sensitive writing systems like Urdu Nasta'leeq. This project presents a scientific study of Nasta'leeq orthography and models it using OTF specification.

Previously there was no freely available open type font for Urdu available in the market which could write in Nasta'leeq. Development of this font has enabled the users to publish electronically in Urdu and thus reach out to the extensive readership across the world. This font is freely downloadable from www.crup.org and is enabled within the existing applications and web browsers (which support OTF). This enables free and extensive access to wide range of computer users.

Project Design and Implementation

In this project the project design evolved in the form of physical model for Nafees Nasta'leeq. The final output of this phase was a document that covered details of all shapes to be stored, their substitution and positioning grammar, nuqta placement and kerning rules. It also included stepwise description of all processes involved in font development. The phases involved in the project are:

1. Domain Understanding and literature review

To understand previous work in this field we proceeded in two directions. Firstly, tool, technologies and standards defined for font development were studied in depth. Tools that were needed for this project were also indicated and later bought. Secondly, currently available Urdu fonts and application supporting Urdu were gathered and tested upon.

2. Requirements gathering

The major part of the research was to extract the information and requirements from the calligraphers and then model the system in accordance to those requirements. The calligraphers follow intuitive rules while writing in Nasta'leeq. These intuitive rules were explained by the calligrapher during sessions with him. The various shapes of the characters at a particular context were also documented.

3. Prototype development

As a prototype we developed an OpenType font for Naskh which is a particular writing style of Urdu with simpler substitution and positioning rules in comparison to Nasta'leeq writing style. The development life cycle of this prototype was shorter though similar to that of Nasta'leeq.

4. Design phase

The project design evolved in the form of physical model for Nafees Nasta'leeq. It covered in detail the work done at TTF level and the effort carried out at OTF level. In this phase individual glyphs which form building blocks of our font were listed and finalized. Procedures to be adopted to convert these glyphs from image to splines (standard mathematical form for glyph outline) were defined. Similarly work to be done at True Type level was also decided, such as space width and

inter ligature space etc. Lastly, contextual rules for substitution and positioning were thoroughly indicated.

5. Implementation

Implementation phase can be broadly divided into three sub phases:

Image Processing

Modules in this phase includes: acquisition of Nasta'leeq glyphs and ligatures, scanning of Nasta'leeq glyphs, segment join verification and ligature shape refinement, thresholding of Nasta'leeq glyphs (monochrome), and vectorization of Nasta'leeq glyph outlines

TrueType Level Work

Steps involved in this phase include: glyph size for Nasta'leeq font, positioning of glyphs at TTF level, joining of glyphs at TTF level, including missed shapes, glyph ordering and TTF file organization.

OpenType Level Work

Steps involved in this phase include: Unicodes assignment, glyph grouping, implementing substitution, positioning, cursive positioning, mark positioning, and kerning rules

6. Testing

Robust testing was done at three levels. Firstly, all valid ligatures of Urdu were tested. Secondly, written texts of Urdu from various fields were tested. Thirdly, 1.7 M words were taken from frequency analysis of Urdu done by CRULP and tested upon.

7. Release

At the end of this exercise Nafees Nasta'leeq font was released. This font is freely downloadable from www.crupl.org

Other activates supported under this project are listed below:

- **Built a prototype of Nafees Naskh**

As a prototype we developed an OpenType font for Naskh which is a particular writing style of Urdu with simpler substitution and positioning rules in comparison to Nasta'leeq writing style. The development life cycle of this prototype was shorter though similar to that of Nasta'leeq.

At the end of this exercise Nafees Naskh - an OTF font was released. This font is freely downloadable from www.crupl.org . CRULP's own Urdu website has also been developed in this font. Also a local magazine (Akbar-e-Urdu June 2003) by National Language Authority contains research papers published in this font.

In addition, this font was extended to include aerabs (vowel marks) through Microsoft Pakistan's Localization R&D Grant to CRULP. Thus this work which started as a prototype took form of a full fledge product and acted as a foundation stone to help boost font industry in Pakistan.

- **Nasta'leeq's lectures**

The calligraphers follow intuitive rules while writing in Nasta'leeq passed down through generations of calligraphers. These intuitive rules were explained by the calligrapher Jamil-ur-Rehman during regular sessions. Duration of this activity was four months. These rules were quantitatively examined and published in 'Orthographic Analysis of Nasta'leeq Writing Style for Urdu' (submitted in December 2002).

- **Font Seminar**
An educational seminar regarding font development was organized by Technology resource Mobilization Unit (TReMU) at Arid Agriculture University, Rawalpindi, Pakistan. The purpose of this workshop was to create awareness about life cycle of font development. The presenters were all from CRULP trained through this grant. This workshop was attended by around 200 attendees.

Project Outputs and Dissemination

The major outputs of this project are listed below.

1. A comprehensive document on Logical Model of Nasta'leeq titled 'Orthographic Analysis of Nasta'leeq Writing Style for Urdu'.
2. A document with details of context-dependent shapes of each letter and their examples, titled 'Context-Dependent Shape Inventory'.
3. A comprehensive document describing the design and implementation details of Nasta'leeq titled 'Physical Model of Nafees Nasta'leeq Writing Style for Urdu'.
4. Nafees Nasta'leeq Font
5. Release note

Dissemination efforts relevant to this project include:

1. The above has been made available for free download at the website of CRULP (www.crupl.org).
2. The developed font has been advertised in daily 'Jang' (the most widely circulated Urdu news paper in Pakistan) on 13th August, 2003.
3. This font has also been advertised at relevant mailing groups dedicated to font development such as MSVOLTCommunity, PakType, OpenType etc.
4. Efforts required for similar activity has been disseminated through an educational seminar regarding font development which was organized by Technology resource Mobilization Unit (TReMU) at Arid Agriculture University, Rawalpindi, Pakistan (as discussed above).

Capacity Building

After successfully going through the font development process there is an increase in research skills of the researchers and interneers involved. The contributions of the researchers include organizing free font training sessions at font seminar to provide an insight into font making.

In addition, the grant has helped in acquiring new equipment such as printer, personal computers and scanner. Specialized software required for font development has also been purchased. Thus, the software obtained and experience gained through this project can be utilized by the research team to work on other computationally unexplored scripts. Also this grant has helped in strengthening CRULP's capacity for analysis and research. Through this grant CRULP has matured into a growing center of excellence in computation of Urdu and other regional languages.

Project Management

Center for Research in Urdu Language Processing (CRULP) has capacity to plan, manage, monitor and deliver commercial strength software. Team Leader and his team have all worked in professional development environment for many years before joining the research team at CRULP. Therefore, most of the problems were identified in the analysis phase and remedial measures were taken to address them before they occurred. As an example, OTF technology was thoroughly tested through prototyping Nafees Naskh font. This exercise also served the dual purpose of training the student team the technical details of the new font formalism. This proved to be a useful exercise for actual design and development of Nafees Nasta'leeq font and save a lot of time.

The project was managed in separate planning, analysis, design, implementation and testing cycles, following conventional software engineering practices. Extensive testing was done to identify three levels of bugs. All level-one bugs were removed. Though other level bugs were also removed, some level-two and some level-three bugs were left open because the team ran out of space within the OTF formalism to write any further rules.

The Beta version of font was released on 14th August, 2003 and is currently being maintained to remove any additional bugs being reported by end users (which have been very minimal).

CRULP Team has been very satisfied with the very patient AMIC administration involved in the project. They have been very understanding and accessible over email. We had sent two budget revisions over the 18 month period and they have been very quick and positive in their response. There was a query in the beginning of the project about the effectiveness and intentions of the project by a private vendor from Pakistan. AMIC administration handled it very discretely. Finally, the payments were made very efficiently after submission of relevant documents. There was some miscommunication between AMIC and CRULP in the first payment, where we could not trace the payment for some time. It was eventually credited to our account and AMIC was very helpful in tracing the money transferred.

Though we were not directly associated with IDRC through the project, the first contact was made through the invitation which was sent to the Project Leader to attend the PAN Asian conference in Vientiane. The conference was a wonderful opportunity to network with other PAN partners and to see other problems and solutions. The conference was extremely well managed and the operations for the conference were smoothly handled from our perspective.

Impact

One of the major stumbling blocks to progress in the field of information and communications technologies (ICTs) in countries like Pakistan is the inability of ICTs in general to provide solutions in local languages. Development of various Urdu fonts will in turn lead to development of Urdu websites and user interfaces. Thus, foreign language barrier for formal correspondence has been greatly reduced.

Development of Nafees Naskh and Nasta'leeq font is an innovative new ICT utility of relevance to concrete local situations. This has acted as a catalyst to further analyze and develop fonts in regional languages other than Urdu. It can also provide an opening for other donors to consider their own role in the promotion of Urdu computation. In this regard CRULP has already won a grant from Government of Pakistan to develop a software system for machine translation.

The overall impact of the project has still to be realized. The font has been made, but still a lot of people do not know how to use it. It is easy for an expert user to download font, read the

Readme file and install the font. However, a novice user finds it hard to do. Though some marketing has been done, just marketing is not sufficient. A significant outreach program should be designed which should train the end users across the country. However, for the current project such work was not envisioned. However, such work should be followed soon to create the requisite impact.

Overall Assessment

Nafees Nasta'leeq has been a truly exhilarating project. Everything worked out well. The results can be attributed to the hard work of the core team (the funded students at CRULP) who spent long hours and days and eventually delivered. The credit also goes to our calligraphers who worked with a great deal of understanding and patience with a team of computer scientists, former being a different breed of professionals. Even worlds apart, there did not seem to be any significant communication problems between them. This project has built capacity at CRULP to develop fonts. This capacity is now being shared by other organizations, for example, CRULP team did a free training on Urdu font development at Arid Agriculture University.

This project has also been extremely successful in documenting logical and physical model for Nasta'leeq script. Research documents in this regard are attached with this report. Such detailed level of work has not been documented on Nasta'leeq writing style to date.

The font development work has also tested the limits of OTF formalism, which is currently being extended to give space for more rules to realize complex fonts like Nasta'leeq more truly.

This has also enabled 60 million Urdu users to develop web pages and publish and easily access content using the Internet.

Recommendations

Font making is an arduous and tiring process, requiring hours of development and testing "to make it look just right." Unprofessional development can result in very unaesthetic fonts, which make them useless, as people using fonts (even naïve computer users) can easily tell good from bad as they are very familiar with their writing systems. Initially we wanted to make the font totally open source. However, though we are distributing the font for free even now (as had been initially committed) hard work has developed a realization that it may not be in the best interest of the work to make it open source. If any user can start experimenting with the font and start making their own version, it will create multiple versions, most of them not very aesthetically pleasing, and thus create a lot of confusion for the end users. Thus, we feel that professional fonts should not be open source.

Also, related to this matter are commercialization opportunities of such products. We were contacted by BBC, UK at the release of the font. They are interested in using it for BBC Urdu website, and wanted to purchase rights so that they can modify and further improve upon the work. However, as the IPR issues are detailed in the grants, it is difficult to assess how such matters may be handled. Administering agencies (AMIC, IDRC, APDIP UNDP, APNIC) should also look into advising and guiding their partners in relevant IPR issues, and making the agreement much more detailed in this area.