

平成 28 年 6 月 21 日現在

機関番号：34514

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25370390

研究課題名(和文)古ロシア語文献デジタル化の為に、初期キリル文字を含むウェブフォントの開発について

研究課題名(英文)on making a Unicode fonts including "Cyrillic Extended" characters for digitization of old Russian literature

研究代表者

酒井 純 (SAKAI, Jun)

神戸親和女子大学・文学部・准教授

研究者番号：20390049

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：この研究では、初期キリル文字を含む2つのフォントの作成を最終年度までにすべて終了した。また他の開発者によるGNU Free Fontについて、そのライセンスおよび収録する文字が本研究の目的と合致することが確認された。これらにより本研究の目的である、古ロシア語で用いられるキリル文字を含むフリーライセンスのウェブフォントを作成することができたと考えている。

またこれらのフォントが完成したことで、これまで作成に協力してきたコンコーダンスについても、Unicodeに合致した形で利用可能になった。実際8つの年代記についてUnicodeへの対応の作業をして、上記フォントと併せてウェブ上で公開している。

研究成果の概要(英文)：In this study, all completed the creation of two fonts, including the Early Cyrillic alphabet characters in the last fiscal year. As for the software license and including characters of GNU Free Font created by other developers, it was confirmed that matches the purpose of this study. By these, the purpose of this study, to create a free licenses web fonts that contains all Cyrillic characters used in the Old Russian, is considered to have achieved. These fonts were to be able to use the concordances that we created in Unicode. And character code of eight Chronicles was converted to Unicode, it has been published on the web in conjunction with the fonts created.

研究分野：コンピューター言語学

キーワード：初期キリル文字 フォント ウェブフォント Unicode Version 8.0 基本キリル文字 補助キリル文字
拡張キリル文字A 拡張キリル文字B

1. 研究開始当初の背景

この研究開始前に研究協力として関わっていた、ロシア諸年代記のコンコードランスを作成する作業の中で、現在のロシア語などでは使われないキリル文字成立当初の文字（以下初期キリル文字と表記）を含むフォントを作成していた（「」など）。

これらの研究の過程で、コンピューターの OS の世代交代や Unicode の普及により、コンコードランスも Unicode にしたがって作成できる環境がそろってきたが、求める文字が Unicode にもない場合には、外字領域を使って対応をしてきた。このような状況の中で、Unicode5.1 が発表され、これに従ったフォントであれば、これまで作成した初期キリル文字はすべて対応できることとなった。しかしながら実際には Unicode5.1 に含まれる初期キリル文字をすべて含むフォントは、現在まで Windows に収録されていない。

このため、Unicode5.1 以降に対応したキリル文字フォントを作成し、True Type Font および Web Font として、ライセンスフリーの形で公表することができれば、コンコードランスでの初期キリル文字の問題は一通りの解決になると考えた。

2. 研究の目的

本研究の目的は、初期キリル文字を含むフォントを Unicode に沿う形で作成することで、これまで古ロシア語の文献データでできなかった検索や文字列処理の問題を解決する。また、このフォントをフリーのウェブフォントとして公開することにより、ロシア語系フォントの著作権問題を解決するとともに、インターネット上での利用を促進する。

これにあわせて、これまでに作成されたロシア諸年代記コンコードランスのデータを Unicode に対応したデータに変換する。これらにより、これまでのデータの標準化を図り、利便性を向上するとともにアーカイブとしての価値の向上を図るものである。

3. 研究の方法

この研究を進める方法としては、まず Unicode でのキリル文字の収録状況の経過をたどるとともに、どの Version に合わせたフォントの作成が妥当かを検討する。また同時に、これまでの MS-Windows に標準搭載されたフォントの、Unicode 対応の状況を確認する。

次に、初期キリル文字を含む Preslav フォントの True Type 版を作成する。フォントの書体については、これまでと同様に各コンコードランスの写本を参考に、文字のデザインを決定した。また True Type Font の作成後に、変換ツールを用いてウェブフォントへの変換を行う。

また、フォントの作成が進んだ研究 2 年目からは、各種コンコードランスなどの古ロシア語文献の改定作業を始める。この改定作業は、

同時に上記フォントのバランスチェックを兼ねており、紙面上および画面上での可読性と画面表示状態の確認を行った。

4. 研究成果

(1) まず、キリル文字の Unicode への収録状況の経過を明らかにすることができた。Unicode5.0 以前は、基本キリル文字および補助キリル文字の U+0400~U+0513 までが収録されており(U+0487 をのぞく)、このバージョンまで収録文字は大きく変更されてはいなかった。初期キリル文字が大量に追加されたのが Unicode5.1 で、ここでは拡張キリル文字 A 領域の U+2DE0~U+2DFF、および拡張キリル文字 B 領域の U+A640~U+A69F までが追加されている。これ以降のバージョンでは、少しずつ文字が追加されており、2015 年に発表された Unicode8.0 で、基本キリル文字、補助キリル文字、拡張キリル文字 A、拡張キリル文字 B のすべての文字（432 文字）が満たされた。これらの更新状況をまとめたものが、次の表 1 である。

バージョン	キリル文字関連の変更点
Unicode 5.0	U+0400~U+0486 U+0488~U+0513
Unicode 5.1	U+0487 追加 U+0514~U+0523 追加 Cyrillic Extended-A 表追加 (U+2DE0~U+2DFF) Cyrillic Extended-B 表追加 (U+A640~U+A69F)
Unicode 5.2	U+0524, U+0525 追加
Unicode 6.0	U+0526, U+0527 追加 U+A660, U+A661 追加
Unicode 6.1	U+A674~U+A67B 追加 U+A69F 追加
Unicode 6.2	-
Unicode 6.3	-
Unicode 7.0	U+0528~U+052F 追加 U+A698~U+A69D 追加
Unicode 8.0	U+A69E 追加

表 1. Unicode 各バージョンで追加された文字一覧

(2) 一方、マイクロソフト社の Windows OS に標準搭載されている代表的セリフ体フォントである Times New Roman フォントの、初期キリル文字への対応についてである。Windows XP では、このフォントの Unicode への対応は Unicode1.1.5 相当であった。そして基本キリル文字および補助キリル文字について、Windows 7 では Unicode5.0 相当、Windows 8 では Unicode6.0 相当、Windows 10 では Unicode 7.0 相当であることが確認された。また拡張キリル文字 A、拡張キリル文字 B については、Windows 10 でも対応されていない。

この状況をまとめたものが次の表 2 である。ただし、Windows 8 以降標準ユーザーインターフェイス用フォントとして導入された Segoe UI フォントでは、Windows 8 で Unicode6.0 に対応し、Windows 10 では Unicode7.0 に対応し、拡張キリル文字 A、拡張キリル文字 B の文字も含まれていることが確認された。

バージョン	収録されている文字
Windows XP	基本 Unicode 1.1.5 相当 補助 Unicode 1.1.5 相当
Windows 7	基本 Unicode 5.0 相当 補助 Unicode 5.0 相当 拡張 A なし 拡張 B なし
Windows 8	基本 Unicode 6.0 相当 補助 Unicode 6.0 相当 拡張 A なし 拡張 B なし
Windows 10	基本 Unicode 7.0 相当 補助 Unicode 7.0 相当 拡張 A なし 拡張 B なし

表 2. MS-Windows の各バージョンの Time New Roman フォント Unicode 対応状況

一方、研究開始当初は考慮に入れていなかった、いわゆるモバイル OS (Android および iOS) については、酒井 2014 時点で代表的であった Android 4.4 および iOS8.1 のブラウザ上で、ともに Unicode 5.0 相当であることが確認されている。

また、Windows 7 以降の上記すべての OS に標準搭載されるウェブブラウザ について、ウェブフォントを用いることで、キリル文字に関するすべてのフォントが表示可能であることを確認した。

(3) これらの状況を確認したうえで、この研究でのフォント作成の方針を以下のように決定した。

文字の収録範囲としては、基本ラテン文字領域の 94 文字、基本キリル文字領域の 256 文字、補助キリル文字の 48 文字、拡張キリル文字 A 領域の 32 文字、拡張キリル文字 B 領域の 95 文字 (のちに U+A69E 「」を加えた 96 文字) の、全 525 文字 (後に 526 文字) とする。フォントのデザインは、オリジナルのセリフ体とする。

ソフトウェアとしてのライセンス形態としては、GNU GPL font exception clause (以下 と記述) とする。

(4) 実際のフォントの作成作業としては、基礎となるフォントのデザインを部品単位で決定し、これを組み立てることで各文字を作っていた。作成作業は武蔵システム社の

TTEdit 上でを行い、酒井 2014 では、初期キリル文字を含む 525 文字の、Unicode7.0 に対応した Preslav フォント (Version 1.0) を作成した。

また、Izhitsa フォントを利用して初期キリル文字の表示を行っているコンコードダンスの対応のために、基本ラテン文字領域に初期キリル文字を割り当てた、Dobrogea フォントも作成している。

どちらのフォントについても、まず True Type 形式のフォントを作成したうえで、woff コンバーター (同武蔵システム社) を使って woff 形式のウェブフォントに変換している。そしてこれらのフォントは、ライセンスを として、ウェブ上で公開している。

(5) 今回作成したフォント以外でも、初期キリル文字を含み、コンコードダンスに利用が可能と考えられるフリーフォントとして、GNU Free Font (別名 Free UCS Outline Fonts) が見つかっている。このフォントは、GNU のライセンスに従ったフリーフォントで、Open Type Font、True Type Font、WOFF Font の各形式に対応しており、キリル文字についても Unicode 6.0 に準拠し、拡張キリル文字 A/B についてもすべて収録されている。このフォントは、初期キリル文字以外にも多くの文字を収録しており、Preslav の優位性としては、2016 年 6 月現在、キリル文字の領域が Unicode 8.0 である点のみである。

GNU Free Font は 2016 年 6 月現在も Unicode6.0 相当が発表された段階で止まっており、現在のところこの点が優位に立っているといえるが、今後のこのフォントの開発状況によっては、Preslav の開発を Version 1.1 までで終了させてもよいかと考えている。

(6) 2015 年度の研究では、まず Preslav フォントのバージョンアップを行った。これは、Unicode 8.0 で、拡張キリル文字 B 領域に 1 文字追加があったためである (U+A69E 「」)。これにより Preslav フォントは Version 1.1 となり、GNU Free Font の今後の利用の可能性からもここまで開発を終了することとした。

次に 2014 年度から取り組んでいたロシア諸年代記の Unicode 対応についてであるが、Unicode を基本として作成された次の 8 つの年代記については Unicode 対応作業を終え、静的コンコードダンスの形ではあるがウェブ上での公開を行うこととなった。

- モスクワ年代記集成 (XV 世紀末) コンコードダンス
- ノヴゴロド第 4 年代記コンコードダンス
- イパチー年代記コンコードダンス
- ソフィア第一年代記コンコードダンス
- プスコフ諸年代記コンコードダンス
- ロシア原初年代記コンコードダンス

- ・ラジヴィル年代記コンコーダンス
- ・トロイツア年代記コンコーダンス

今回作成したウェブフォントなどを用いることで、これらすべてのコンコーダンスで初期キリル文字を標準ブラウザにて表示することが可能となり、これは各バージョンの Windows および Android、iOS 上でも表示の確認を行っている。

また、これ以外の Tex を用いて作成されたコンコーダンスについては、今回の研究期間内には Unicode 対応を終えることはできなかったが、今後も継続して改訂作業を行い公開してゆく方針である。

(7) 今回の研究では、まず現在の代表的な OS であるマイクロソフト社の Windows シリーズでの初期キリル文字への対応、および Unicode 対応の状況をまとめることができた。

この状況を踏まえた上で、今回の研究期間内に、第 1 の目標である初期キリル文字などを含む Unicode 8.0 に対応したフリーのウェブフォントを作成することができた。このことにより、ロシア諸年代記のコンコーダンスでの表示問題、およびフォントのライセンス問題を解決できた。

また、第 2 の目標としてきたコンコーダンスの Unicode 対応の改訂作業についても、前述の 8 つのコンコーダンスについてその作業を終えて、公開することができる段階まで来た。

今回の研究の成果であるウェブフォントなどを用いることで、ロシア諸年代記を含む古ロシア語文献のデジタル化に寄与できたのではないかと考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

酒井純、「拡張キリル文字群を含む Unicode フォントの作成について」神戸親和女子大学言語文化研究、査読無、第 9 巻、2015 年、pp.1-20

〔その他〕

ホームページ

キリル文字フォント

<http://www.scll6.info/fonts/>

ロシア諸年代記コンコーダンス

<http://www.scll6.info/concordance/>

6. 研究組織

(1) 研究代表者

酒井 純 (SAKAI, Jun)

神戸親和女子大学・文学部・総合文化学科・准教授

研究者番号：20390049