

# **Digital Preservation and Access to Cultural and Scientific Heritage: Presentation of the KT-DigiCult-Bg project**

**Dr. Nikola IKONOMOV**

**Institute for Bulgarian Language, Bulgarian Academy of  
Sciences, Sofia, Bulgaria**

**[nikonov@ibl.bas.bg](mailto:nikonov@ibl.bas.bg)**

**Dr. Milena DOBREVA**

**Institute of Mathematics and Informatics, Bulgarian  
Academy of Sciences, Sofia, Bulgaria**

**[dobreva@ufal.ms.mff.cuni.cz](mailto:dobreva@ufal.ms.mff.cuni.cz)**

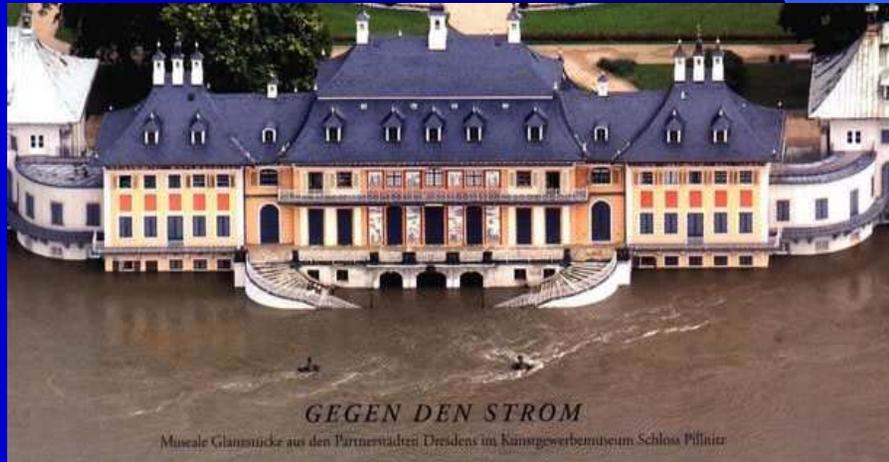
# Contents

- Introduction
- Previous Bulgarian experience
- The KT-DigiCult-Bg project
- Conclusions

# Introduction

- Digitisation of cultural and scientific heritage is among the EU priorities for two basic reasons:
  - necessity for heritage preservation.
  - new understanding of access to the heritage

# Heritage Preservation



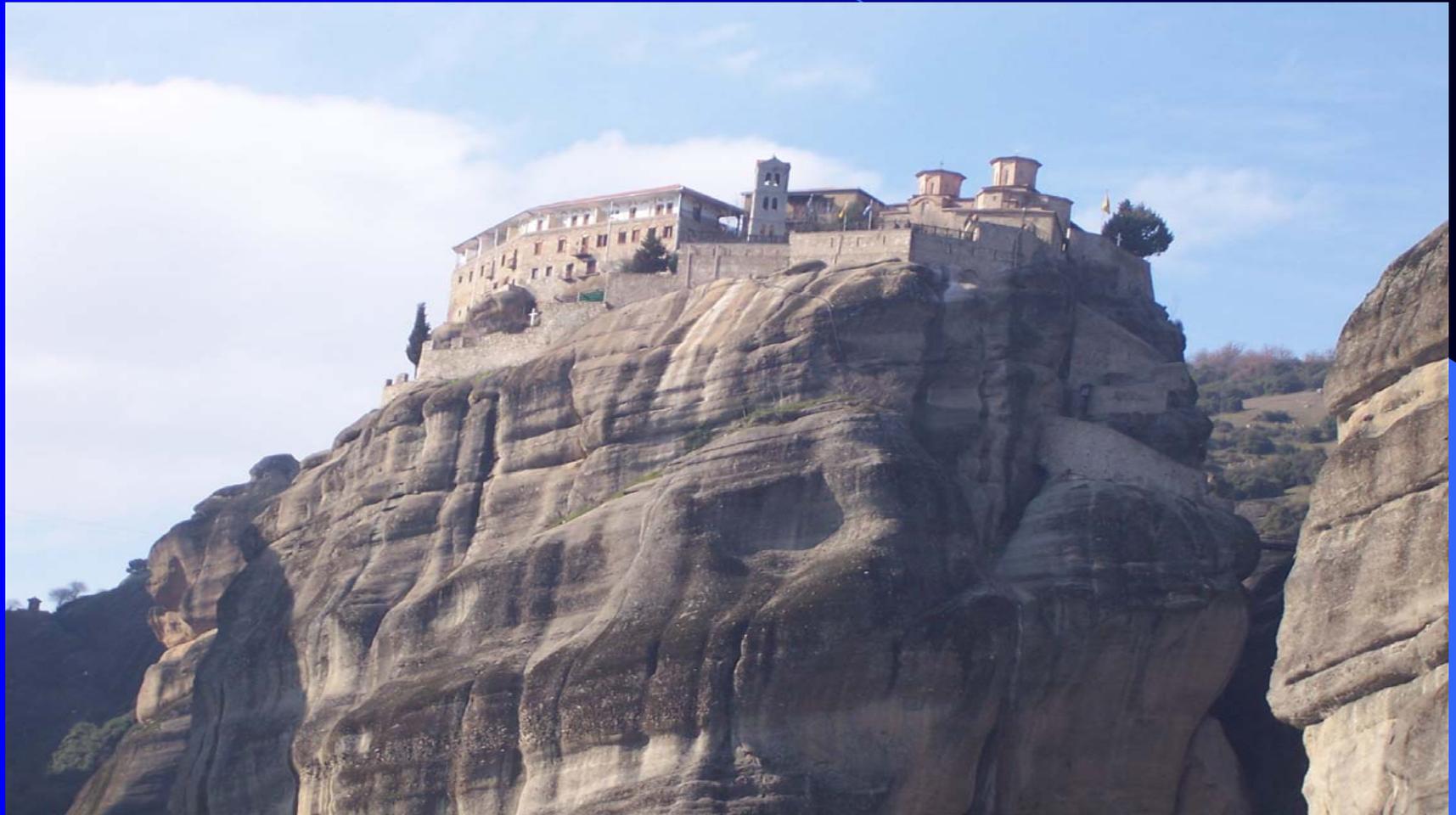
# Some ideas about preservation



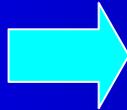
# Cultural and scientific heritage preservation

- Preservation means to prepare physical copies of the most valuable products of the human intellect and culture in order to save them for the future generations.

# Some aspects of the access to the heritage



# From restricted community to the general public

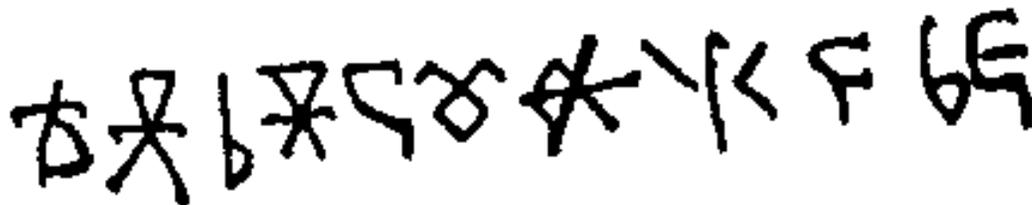




- South-Eastern Europe is underrepresented in the cultural space

# Bulgarian heritage assets

- 8500 Christian manuscripts + 4000 Islamic
- More than 35000 early printed books
- Third largest collection of epigraphic inscriptions in Latin and Ancient Greek in the world
- Old Bulgarian runic inscriptions
- ... and numerous other



Old Bulgarian runic inscriptions, showing a sequence of stylized, angular characters.

# Previous experience – 1

- **Cataloguing:** ISIS (15 years ago), Repertorium (300), MASTER (30), XEditMan (850)
- **Electronic publications on CD-ROMs:** Balkan manuscripts, The Holy Koran, Bulgarian Iconography; The Treasury of Lysimachos

# Previous experience – 2

- **Presentation of images** (An Isolated example of digitized images not supplied with texts): The Enina Apostle
  - 11 century (the oldest Bulgarian manuscript stored in Bulgaria)
  - 39 folios survived in a lime-pit, badly damaged

# The Enina Apostle

Microsoft Internet Explorer window titled "Thumbnails--The Enina Apostle". The address bar shows the URL: [http://www.rch.uky.edu/~kiernan/cgi-bin/enina\\_ebind.cgi/enina?thumbs=17;target=1](http://www.rch.uky.edu/~kiernan/cgi-bin/enina_ebind.cgi/enina?thumbs=17;target=1).

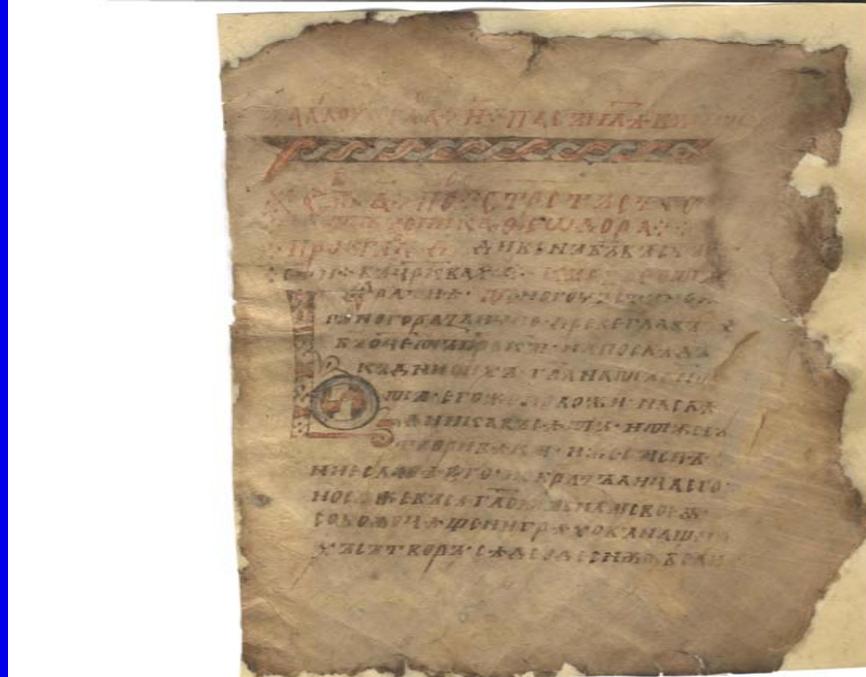
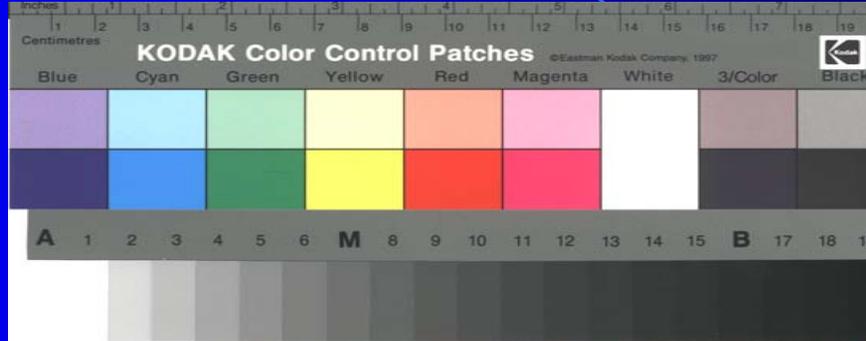
The main content area displays a grid of eight thumbnail images of manuscript pages, arranged in two rows of four. The top row images are labeled below as:

- fol. 9r
- fol. 9v
- fol. 10r
- fol. 10v

The bottom row contains four additional thumbnail images of manuscript pages, which are not individually labeled. The images show fragments of parchment with handwritten text in a medieval script.

The browser's status bar at the bottom indicates "Done" and "Internet". The taskbar shows the Start button and several application icons, with the system clock displaying "09:59".

# The Enina Apostle



# Previous experience - 3



- 3D models of immovable heritage (Bojana church- 13<sup>th</sup> century )
- Digitization and Restoration of Audio Archives

# The KT-DigiCult-Bg project

- The project *Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria* (KT-DigiCULT-BG) is a four year project supported by the Marie Curie programme.
- It is coordinated by the Institute of Mathematics and Informatics (Bulgaria).
- Project partners are: Det Arnamagnæanske Institut (Københavns Universitet, Denmark), Trinity College (Dublin, Ireland), Charles University (Prague, Czech Republic), and the Institute of Informatics and Telecommunications, National Centre for Scientific Research “Demokritos” (Athens, Greece)

# Project specifics

- The project is a typical Transfer of Knowledge action . It is structured as two-phase development scheme:
  - Knowledge acquisition phase
  - Knowledge transfer phase
    - Direct guidance
    - Establishment of a National Digitisation Centre at the Host Institution

# Basic field of work

- General methodology and practical setting for digitisation of cultural and scientific heritage.
- Digitisation of mediaeval manuscripts (incl. digital imaging, cataloguing, text representation, electronic publishing).
- Digitisation of mathematical texts and building digital mathematical library of works of Bulgarian mathematicians.
- Virtual reality applications for presentation of immovable cultural heritage.
- Audio archives: methods for digitisation and restoration.
- Application of quantitative methods for the study of data related to the cultural heritage.
- Applications of edutainment to cultural heritage studies.)

# Some results after the first project year

- IMI (the host) started the digitisation of its scientific heritage.

**АЛГОРИТМ ДЛЯ НАХОЖДЕНИЯ ВСЕХ ИЗОМОРФИЗМОВ ДВУХ ГРАФОВ**

АТАНАС А. РАДЕНСКИ

В работе предлагается эффективный алгоритм для нахождения всех изоморфизмов двух неориентированных или ориентированных графов. Время, необходимое для нахождения одного изоморфизма двух  $n$ -вершинных графов, пропорционально  $n^6$ . Алгоритм можно использовать, например, для решения некоторых задач химии и теории электрических цепей.

**1. Введение.** Пусть  $L$  — неориентированный связный униграф без петель с вершинами  $x_1, x_2, \dots, x_n$ . Элементы матрицы смежности  $A_L = (a_{ij})$  графа  $L$  определяются следующим образом:

- если вершины  $x_i$  и  $x_j$  смежны, то  $a_{ij} = a_{ji} = 1$ ;
- если вершины  $x_i$  и  $x_j$  несмежны, то  $a_{ij} = a_{ji} = 0$  ( $i, j = 1, 2, \dots, n$ )

Отметим, что вершина  $x_i$  несмежна с вершиной  $x_i$ . Матрица смежности  $A_L$  вполне определяет граф  $L$ .

Пусть  $L$  и  $M$  — два графа с матрицами смежности  $(a_{ij})$  и  $(b_{ij})$  и с множествами вершин  $X = \{x_1, x_2, \dots, x_n\}$  и  $Y = \{y_1, y_2, \dots, y_n\}$ . Графы  $L$  и  $M$  называются изоморфными ( $L \cong M$ ), если существует взаимно однозначное отображение  $\varphi$  множества  $X$  на множество  $Y$ , такое, что из равенств  $y_p = \varphi(x_i)$ ,  $y_q = \varphi(x_j)$  следует  $a_{ij} = b_{pq}$  ( $i, j = 1, 2, \dots, n$ ). Теоретически проблема нахождения всех изоморфизмов графов  $L$  и  $M$  решается просто — проверкой всех  $n!$  взаимно однозначных отображений  $\varphi$  множества  $X$  на  $Y$ . Однако практически такое решение неприменимо. Для некоторых специальных графов существуют эффективные решения проблемы изоморфизма [2, 3, 4]. В работе [1] предлагается общий алгоритм для нахождения всех изоморфизмов двух графов, у которого порядок быстродействия —  $n^6$ . Этот алгоритм основан на нахождении длинного списка характеристик. Совокупность списков предложенного вида оказывается инвариантной относительно изоморфизма. В [6] описан один из интереснейших алгоритмов для проверки изоморфности двух графов, в котором используется одно недоказанное предположение.

В настоящей работе предлагается весьма простая характеристика, определяющая граф с точностью до изоморфизма. Эта характеристика позволяет выяснить, изоморфен ли граф  $L$  графу  $M$ , и если они изоморфны, найти все изоморфизмы. Предложенный алгоритм без особых изменений применим и к случаю, когда  $L$  и  $M$  — произвольные униграфы.

**2. Определения.** Подграфом графа  $L$  называется каждый граф  $L'$  с множеством вершин  $X', X' \subseteq X$ , матрица смежности которого получена из  $A_L$  путем удаления всех строк и столбцов с номерами  $i$ , для которых  $x_i \notin X'$  и  $x_i \notin X'$  ( $i = 1, 2, \dots, n$ ), т. е. две вершины из  $X'$  смежны

СЕРДИКА. Българско математическо списание. Том 1, 1975, с. 167—177.

138

ИЗВЕСТИЯ ИМ. АКАДЕМИИ НАУК БЪЛГАРИИ

**Следствие 1.** Пусть  $Y = (y, \infty)$ -компактное подпространство вполне регулярного пространства  $X$ . Тогда для любой теоретико-множественной  $\tau$ -операции  $\psi$  и любого  $n < \aleph^+$  имеем:

$$\psi - \bar{\psi}_n(Y) = \{Y \cap L \mid L \in \psi - \bar{\psi}_n(X)\},$$

$$\psi - U_n(Y) = \{Y \cap L \mid L \in \psi - U_n(X)\},$$

$$\psi - \bar{\psi}_n^*(Y) = \{Y \cap L \mid L \in \psi - \bar{\psi}_n^*(X)\}.$$

**Следствие 2.** Пусть  $Y = (y, \infty)$ -компактное подпространство вполне регулярного пространства  $X$ . Если для некоторой теоретико-множественной  $\tau$ -операции  $\psi$  пространство  $Y$  удовлетворяет свойству  $\psi - H$ , то и  $X$  удовлетворяет свойству  $\psi - H$ .

**Теорема 4.** Пусть  $f: X \rightarrow Y$  — совершенное отображение вполне регулярного пространства  $X$  на вполне регулярное пространство  $Y$  веса  $\tau$ . Если для некоторой слабополнозначной теоретико-множественной  $\tau$ -операции  $\psi$  пространство  $Y$  удовлетворяет свойству  $\psi - H$ , то и пространство  $X$  удовлетворяет свойству  $\psi - H$ .

Полнотой этого факта можно пользоваться при построении боровских сечений для многозначных отображений.

**3. Модификация топологии.** Пусть  $\mathfrak{B}$  — топология на множестве  $X$ . Семейство  $\mathfrak{B} - \mathfrak{B}(\mathfrak{B})$  образует базис некоторой новой топологии  $\mathfrak{B}(\mathfrak{B})$  на множестве  $X$ . Топологию  $\mathfrak{B}(\mathfrak{B})$  назовем  $\mathfrak{B}$ -модификацией топологии  $\mathfrak{B}$ .

**Предложение 1.** Пусть  $\mathfrak{B}$  — класс  $\tau$ -операций,  $\mathfrak{B}_1$  и  $\mathfrak{B}_2$  семейства подмножеств множества  $Z$  и  $\mathfrak{B}_3 = \psi(\mathfrak{B})$  для некоторой операции  $\psi \in \mathfrak{B}$ . Если для любой операции  $\psi \in \mathfrak{B}$  семейство  $\psi(\mathfrak{B}_1)$  не инвариантно относительно дополнения, то и семейство  $\psi(\mathfrak{B}_2)$  не инвариантно относительно дополнения для каждой операции  $\psi \in \mathfrak{B}$ .

**Лемма 1.** Пусть  $\psi \in \mathfrak{B}$ . Положим  $\psi^* = \psi(\mathfrak{B})$ . Тогда  $\psi(\mathfrak{B}_1^*) = \psi(\psi^*(\mathfrak{B}_1)) = \psi^*(\mathfrak{B}_1)$ . Поскольку  $\psi^*$  — семейство  $\psi(\mathfrak{B})$  не инвариантно относительно дополнения, Предложение доказано.

Положим  $\mathfrak{B}_n = \mathfrak{B} - \mathfrak{B}(\mathfrak{B} - \mathfrak{B})$ , где  $\mathfrak{B}(\mathfrak{B})$  — топологическое пространство веса  $\tau$  и  $\mathfrak{B} \in \mathfrak{B}$ , то  $\mathfrak{B}(\mathfrak{B}) = \psi_n(\mathfrak{B})$ .

**Следствие 3.** Пусть  $(X, \mathfrak{B})$  — вполне регулярное пространство веса  $\tau$  и  $2^m \leq \tau$ . Если для любой теоретико-множественной  $\tau$ -операции  $\psi$  пространство  $(X, \mathfrak{B}(\mathfrak{B}))$  удовлетворяет свойству  $\psi - C$ , то и пространство  $(X, \mathfrak{B}(\mathfrak{B}))$  удовлетворяет свойству  $\psi - C$  для любой теоретико-множественной  $\tau$ -операции  $\psi$ .

В множестве  $J$  рассмотрим семейство подмножеств  $N_i$ , где:

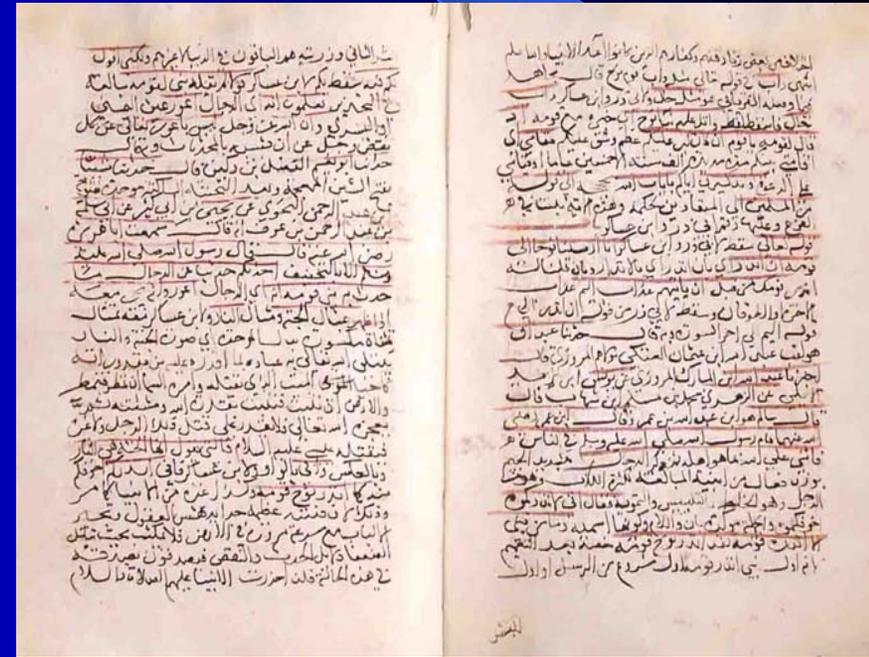
- 1) если  $y \in N_i$ , то  $|x| \leq \tau$ ;
- 2) если  $y_1, y_2 \in N_i$  и  $y_1 \neq y_2$ , то  $y_1 \cap y_2 = \emptyset$ ;
- 3)  $y \in N_i$ ;
- 4) если  $m < \tau$ , то в  $N_i$  имеется ровно  $\tau$  различных элементов мощности  $m$ .

Теперь положим  $K = \{y \subseteq J, y \subseteq N_i \text{ для некоторого } y \in N_i\}$ . Положим  $\sigma = \psi(K)$ . Операция  $\psi$ , являясь положительной\*. Пусть  $P_i$  является операцией пересечения менее  $\tau$  элементов. Тогда  $\psi_i = \psi - S(P_i)$ .

\* — операция  $\psi$  называется положительной, если для любых множеств  $\{E_1, \dots, E_n\}$ ,  $\{A_1, \dots, A_n\}$  из  $E_1 \times \dots \times E_n$  имеем  $\psi(\{A_i\}) \subseteq \psi(\{E_i\})$ . Положительные теоретико-множественные  $\tau$ -операции называются  $\tau$ - $\sigma$ -операциями.

# Some results after the first project year

- A CD- “Sofia. Religious spaces” was prepared and published



## Some results after the first project year

- An Old Cyrillic UNICODE font based on Codex Suprasliensis script (11 c.) was designed

ПРѢЖДЕУЕТЫРЬКАЛАНДЪМАРТА. СНОРѢУЪ

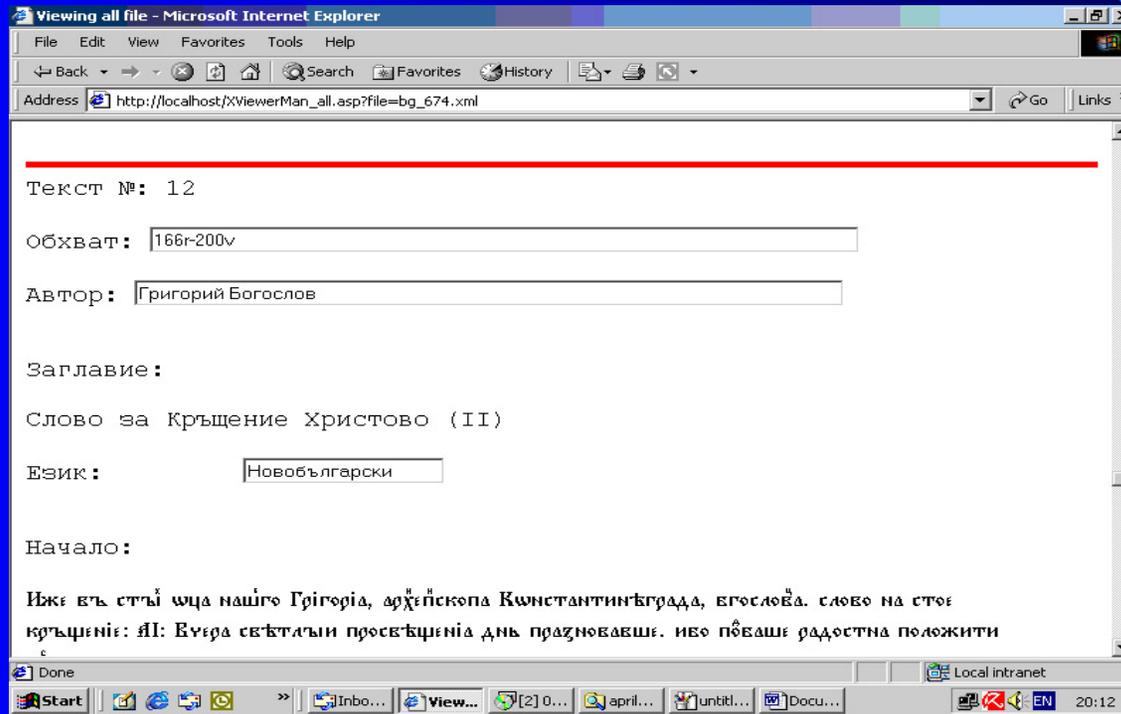
ПРѢЖДЕУЕТЪИРЬКАЛАНДЪМАРТА. СНОРѢУЪ  
ВЪКЪСФЕУРОАРА. ПРѢДАШАЖЕСВОАДОУША  
ГОСПОДЕВН. ПРѢЖДЕУДАНЪМАРТА. ПРНЛН  
КННННСАМОВАЛСТЪНН. НАЖЕЖЕУѢСА РЪСГВХ  
ЖШТОУГОУНАШЕМОУНБѢУНЕСПОУ. Н ВЛА  
ДЪИЧНАШЕМОУИСХСОУ. КМОУЖЕНЕСТЪСЛА  
ВАНДРЪЖАВАНУЕСТЪ. НЪННННПРНМОИВЪ  
ВЪКЪКЪКОМЪ. АМННЪ.



СТААГОВАСНАЪ АРХНЕПЦАКЕСАРНАКАПАДОСНИКЪА.

# Some results after the first project year

- A specialized XML Editor for Manuscript Data for editing and browsing catalogue descriptions of mediaeval manuscripts was developed



# Some results after the first project year

- a methodology of using wavelets for image enhancement and character recognition was elaborated and implemented.

# Conclusion

- Current priorities of EU: IST 2.5.10 (Access to and preservation of cultural and scientific resources)
  - Enriched conceptual representations
  - Advanced access methods
  - Long-term preservation
- We follow the recent developments but still have to do mass digitisation

# Conclusions

- The rich heritage of Bulgaria is still expecting to take its legitimate place in the world cultural space. We hope that through the project we will raise the interest of the scientific and cultural heritage institutions in the country and will push them to start real work.
- The project gives opportunities to broaden the international cooperation in the field of digitisation of cultural and scientific heritage.

# Conclusions

- We highly appreciate the experience of other small countries in Europe, and in particular of institutions from the Czech Republic. The similarity of territory and population, the richness of the cultural and scientific heritage, and the common historical fate in the last century rise comparable problems. The way they are or have been solved in the Czech Republic and the visible progress in the field give promises for a successful Bulgarian digitisation.

Hope I am in time???



Thank you!!!