

Text Search in Document Images Based on Hausdorff Distance Measures

Andrey Andreev¹, Nikolay Kirov²

¹Institute of Mathematics and Informatics, BAS, Sofia

²New Bulgarian University & Institute of Mathematics and Informatics, BAS, Sofia

This research has been partially supported by a Marie Curie Fellowship of the EC programme “Knowledge Transfer for Digitization of Cultural and Scientific Heritage in Bulgaria”.

Introduction

How to find a word in a text document?

When the document is represented as text file, the answer is quite trivial – open the file in any text editor, choose a word and push Find button.

But the task is not so easy when the document is a set of graphic images. This is natural situation when we deal with digitization of cultural and scientific heritage and scanner devices produce files in graphic formats.

Optical character recognition (OCR) is the usual way of conducting text retrieval from scanned document images. OCR software converts text images into a text file, recognizing every letter and mapping it to a number, which is called code. The most often used codes are ASCII (one byte code) or UTF-8 (two bytes code). This technique is well developed and has high accuracy.

But sometimes OCR is a very difficult process requiring dictionaries in the corresponding languages. Often human efforts are needed to correct OCR errors which is quite tedious work. There are some obstacles to successful OCR:

- The quality of page images.
- Language dependency (alphabet and coding, unknown language):
 - dictionaries;
 - old grammar, obsolete words and phrases and idioms;
 - old letters, outside of the coding tables;
 - multi-lingual documents;
- Errors in automatic OCR, human intervention needed.

de la Terre, &c.

fort peu de sens commun.

Quant à la Terre, si vous la rencontrez bonne, ce vous fera un grand avantage, & une grande épargne ; mais rarement en pourrez-vous trouver, où il n'y ait beaucoup à travailler, d'autant que telle paroîtra passablement bonne au dessus, qui étant ouverte de la profondeur d'un fer de Béche seulement, se trouvera Argileuse dessous ; ce fonds est pire aux Arbres que le Tuf, ou la Roche, à cause qu'il s'y rencontre de petites veines où les Racines peuvent s'étendre & profiler, afin de tirer la fraîcheur de plus bas, & prendre quelque nourriture ; mais l'Argileuse ou Terre franche ou rouge, fait comme un plancher qui par sa dureté & densité, ne peut être percé par aucunes Racines, & qui dans les grandes ardeurs de l'Été, em-

meil)uanrΓα, la l'erre,^! vous la rencontrez bonne, ce vous lera un grand avantage, & une grande épargne ; mais rarement en pourrez-vous trouver, on il n'y ait beaucoup à travailler, d'autant que telle paroîtra passablement bonne au dessus, qui étant ouverte de la profondeur d'un fer de Béche seulement, se trouvera Argileuse dessous ; ce fonds est pire aux Arbres que le Tuf, ou la Roche, à cause qu'il s'y rencontre de petites veines où les Racines peuvent s'étendre & profiler, afin de tirer la fraîcheur de plus bas, & prendre quelque nourriture ; mais l'Argileuse ou Terre franche ou rouge, fait comme un plancher qui par sa dureté & densité, ne peut être percé par aucunes Racines, & qui dans les grandes ardeurs de l'Été, em-

We suggest a different approach: instead of applying two steps – OCR and searching in text documents, we will **directly** search words in scanned text documents.

We can organize retrieval of words, similar to a given **pattern word**. The document pages can be represented as binary images in any graphic file format.

The motivation of our work is to choose effective search by simply considering the image similarities. One of the most widespread ideas is to use **Hausdorff type measures** for word image similarity.

Three main steps in the process: segmentation, search and result representation.

- In the segmentation step we create so-called word images – every word is encompassed by a rectangle, which consist of white and black pixels.
- For measuring similarities between word images we use Hausdorff type distances.
- We produce a sequence of words, sorted by values of similarity measure.

The Hausdorff distance (HD) between two point sets A and B is defined as

$$H(A, B) = \max\{h(A, B), h(B, A)\}, \quad (1)$$

where $h(A, B)$ and $h(B, A)$ are co-called directed distances. For the original Hausdorff metrics

$$h(A, B) = \max_{a \in A} d(a, B), \text{ where } d(a, B) = \min_{b \in B} \rho(a, b)$$

is the distance from a point a to the set B , and $\rho(a, b)$ is a point distance.

Euclidean distance: $\rho(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$.

Manhattan distance: $\rho(a, b) = |a_x - b_x| + |a_y - b_y|$.

Infinity norm distance: $\rho(a, b) = \max\{|a_x - b_x|, |a_y - b_y|\}$.

0-1 distance:

$$\rho(a, b) = \begin{cases} 0 & \text{if } a \equiv b \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Huttenlocher *et al.* (1993) proposed the Partial Hausdorff Distance (PHD) for comparing images containing a lot of degradations or occlusions. For directed distance they take the K -th ranked point of A instead of the largest one

$$h_K(A, B) = K_{a \in A}^{th} d(a, B), \quad (3)$$

where $K_{a \in A}^{th}$ denotes the K -th ranked value in the set of distances $\{d(a, B) : a \in A\}$, i.e. for each point of A , the distance to the closest point of B is computed, and then, the points of A are ranked by their respective values to this distance,

$$d(a_1, B) \geq d(a_2, B) \geq \dots \geq d(a_K, B) \geq \dots \geq d(a_{N_A}, B). \quad (4)$$

This HD measure requires one parameter, often represented by $f = K/N_A$ ($0 \leq f \leq 1$). Sim *et al.* claim that a value in the interval $[0.6, 0.8]$ gives good matching results. Note that this measure is not a metric because $h_K(A, A) > 0$!

The idea of José Paumard (1997) is that we do not take into account the L closest neighbors of $a \in A$ in B . So we can define the distance from a point $a \in A$ to the set B as follows

$$d_L(a, B) = L_{b \in B}^{th} \rho(a, b),$$

where $L_{b \in B}^{th}$ denotes the L -th ranked value in the set of distances $\{\rho(a, b) : b \in B\}$ for a given point a of A . Now the directional Censored Hausdorff Distance (CHD) can be defined as

$$h_{K,L}(A, B) = K_{a \in A}^{th} d_L(a, B) = K_{a \in A}^{th} L_{b \in B}^{th} \rho(a, b). \quad (5)$$

Let us set two parameters $\alpha = K/N_A$ and $\beta = L/N_B$ which are relative values with respect to the number of points in the sets A and B . Then the recommended values in for these parameters are $\alpha = 0.1$ and $\beta = 0.01$.

For all three described measures (HD, PHD and CHD), the directed distance can be considered as a choice a representative pair of points (a_0, b_0) , $a_0 \in A$ and $b_0 \in B$ such that the point distance between them $\rho(a_0, b_0)$ is equal to the corresponding directed distance between the sets A and B .

Another approach for measuring similarity between two finite sets in the plane is to calculate a sum of point distances.

Dubuisson and Jain (1994) examined a number of distance measures of Hausdorff type for determination to what extent two point sets on the plane A and B differ. They introduced so-called Modified Hausdorff Distance (MHD) with the following distance measure

$$h_{\text{MHD}}(A, B) = \frac{1}{N_A} \sum_{a \in A} d(a, B) = \frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \rho(a, b). \quad (6)$$

They claim that it suits in best way the problem for object matching. A bit better results were obtained in our examples omitting the coefficient $1/N_A$ in front of the sum. We called this modification Sum Hausdorff Distance (SHD)

$$h_{\text{SHD}}(A, B) = \sum_{a \in A} d(a, B) = \sum_{a \in A} \min_{b \in B} \rho(a, b). \quad (7)$$

In 1999 D.-G. Sim *et al.* described two variants of MHD for elimination of outliers – usually the points of outer noise. Based on robust statistics M-estimation and least trimmed square they introduced M-HD and LTS distances.

The directed distance for M-HD is defined by

$$h_M(A, B) = \frac{1}{N_A} \sum_{a \in A} f(d(a, B)), \quad (8)$$

where the function f is convex and symmetric and has a unique minimum value at zero. One possible function is

$$f(x) = \begin{cases} |x| & \text{if } |x| \leq \tau \\ \tau & \text{if } |x| > \tau \end{cases}$$

This means that we sum the distances $d(a, B)$ which are less than the constant τ and add τ when the distance is greater than τ . The recommended interval of τ is $[3, 5]$. Note that MHD with 0-1 point distance is M-HD for $\tau = 1$.

The second measure is called Least Trimmed Square HD (LTS-HD). The directed distance is

$$h_{\text{LST}}(A, B) = \frac{1}{N_A - K} \sum_{i=K}^{N_A} d(a_i, B), \quad (9)$$

where $K \leq N_A$ and a_1, a_2, \dots, a_{N_A} are points of A for which (4) is valid. Parametrization of the method can be done by a parameter $\alpha = K/N_A$. For comparing noisy binary images the suggested value for this parameter is 0.2.

Following the definition of CHD, we introduce its analogical method based on the sum of point distances. The directed distance is

$$h_{\text{NEW}}(A, B) = \frac{1}{N_A - K} \sum_{i=K}^{N_A} d_L(a_i, B) = \frac{1}{N_A - K} \sum_{i=K}^{N_A} L_{b \in B}^{th} \rho(a, b). \quad (10)$$

We can set again the parameters $\alpha = K/N_A$ and $\beta = L/N_B$ which are relative values with respect to the number of points in the sets A and B .

A new approach to similarity measures

We can consider a linear order of points of A and give a sequence representation: $A = \{a_1, a_2, \dots, a_{N_A}\}$. For every $a_k \in A$ ($k = 1, 2, 3, \dots, N_A$) we can calculate the distances (with respect to a metric ρ in R^2) from a_k to all points in B , i.e.

$$d_k^1 = \min_{b \in B} \rho(a_k, b) = \rho(a_k, b_k^1), \quad d_k^2 = \min_{b \in B \setminus \{b_k^1\}} \rho(a_k, b) = \rho(a_k, b_k^2), \dots,$$

$$d_k^l = \min\{\rho(a_k, b) : b \in B \setminus \{b_k^1, b_k^2, \dots, b_k^{l-1}\}\} = \rho(a_k, b_k^l), \dots,$$

obtaining in such a way a nondecreasing sequence of numbers

$$d_k^1 \leq d_k^2 \leq \dots \leq d_k^l \leq \dots \leq d_k^{N_B}.$$

Carrying out these calculations for every point in A , we define a distance matrix D

$$D = \begin{pmatrix} d_1^1 & d_1^2 & d_1^3 & \dots & d_1^l & \dots & d_1^{N_B} \\ d_2^1 & d_2^2 & d_2^3 & \dots & d_2^l & \dots & d_2^{N_B} \\ d_3^1 & d_3^2 & d_3^3 & \dots & d_3^l & \dots & d_3^{N_B} \\ \dots & \dots & \dots & & \dots & & \dots \\ d_k^1 & d_k^2 & d_k^3 & \dots & d_k^l & \dots & d_k^{N_B} \\ \dots & \dots & \dots & & \dots & & \dots \\ d_{N_A}^1 & d_{N_A}^2 & d_{N_A}^3 & \dots & d_{N_A}^l & \dots & d_{N_A}^{N_B} \end{pmatrix}$$

following arbitrary order of points in A . Later we will choose ordering of rows, corresponding to an order in a column. For definitions of MHD and M-HD we do not need any order

$$h_{\text{MHD}}(A, B) = \frac{1}{N_A} \sum_{i=1}^{N_A} d_i^1, \quad \text{and} \quad h_{\text{M}}(A, B) = \frac{1}{N_A} \sum_{i=1}^{N_A} \min\{d_i^1, \tau\}.$$

For finding the Hausdorff distance in the distance matrix D , we consider the following order (obtained by swapping the rows) with respect to the first column of D

$$h(A, B) = d_1^1 \geq d_2^1 \geq \cdots \geq d_k^1 \geq \cdots \geq d_{N_A}^1.$$

The directed distance for PHD is $h_K(A, B) = d_K^1$. We can calculate LTS-HD summing the part of the first column elements

$$h_{\text{LST}}(A, B) = \frac{1}{N_A - K} \sum_{i=K}^{N_A} d_i^1.$$

We can find CHD directed distance as an element of matrix D swapping the matrix rows in such way that the L -th column is sorted, i.e.

$$d_1^L \geq d_2^L \geq \cdots \geq d_k^L \geq \cdots \geq d_{N_A}^L.$$

Then $h_{K,L}(A, B) = d_K^L$. The directed NEW distance is

$$h_{\text{NEW}} = \frac{1}{N_A - K} \sum_{i=K}^{N_A} d_i^L.$$

Experiments

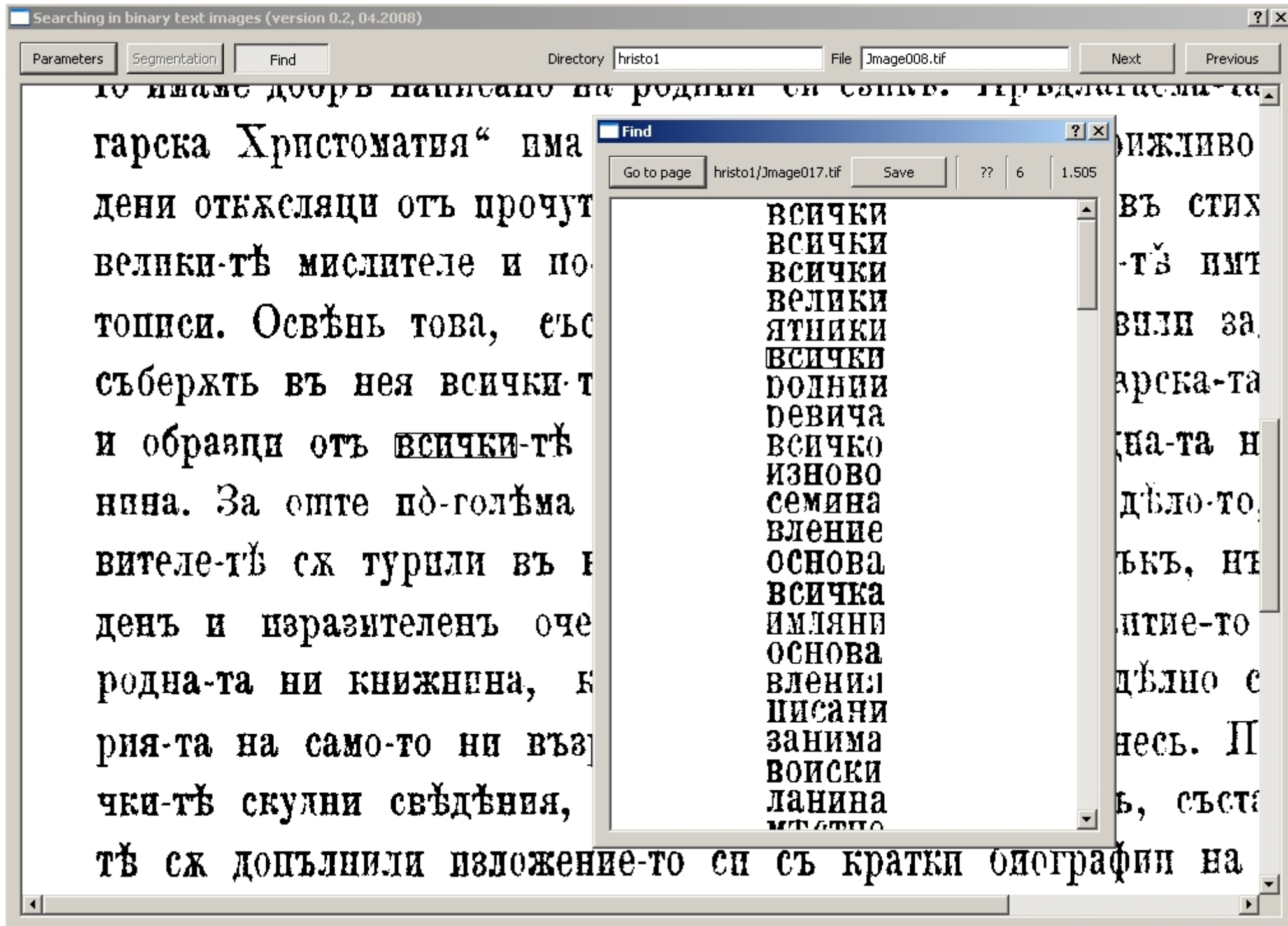
We carried out our experiments using an old book (1884) – Bulgarian Chrestomathy, created by famous Bulgarian writers Ivan Vasov and Konstantin Velichkov. The quality of scanned images are quite bad because this was one of the first books, processing in the digitization center and operators' qualification was not on appropriate level. Many pages have slopes in rows, there are significant variations in gray levels, etc.

There is no text version till now of this book, which may be produced using appropriate OCR software. The first reason is the quality of images. The second reason is the absence of OCR software because the text contains old and abandoned Bulgarian letters. Also spelling and grammar are quite different in modern Bulgarian language.

поетъ, сатирикъ и публицистъ. Първо-то нѣшто, което е издалъ е книжка стихотворения „Басненникъ“ и по-послѣ „Смѣсна Китка“ (Букурештъ 1852 г.), съ които той доби първа-та си извѣстность у насъ, като български писателъ. Отъ 1857 год. се почева неговата многополезна дѣятелность въ борба-та ни съ Гръци-тѣ за черковна независимостъ. Той дохожда въ Цариградъ и издава свои-тѣ „Смѣшни Календари“ сатирически книги, въ които бичува съ единъ искусенъ и ядовитъ сарказмъ пороци-тѣ и недостатки-тѣ на тогавашно-то българско общество, и гръцко-то високо духовенство (1857—1863). На 1863 год. той прѣдприе издаване-то на сатирически вѣстникъ „Гайда,“ който не трая много врѣме. Доста хубави статии все въ полемическо-сатирически духъ, напечата той тамъ. Слѣдъ двѣ години Славейковъ прѣдприе издание-то на политически вѣстникъ „Македония“ (1867—1870). Тамъ при разискване-то на разни въпроси отъ общественъ и черковенъ интересъ Славейковъ се стараше да разбуди народно-то чувство у Македонски-тѣ Българе, които душеше нетърпимо-то влияние на гръкоманство-то и фанариотство-то. Най-послѣ подиръ нѣколко врѣменни спирания и конфискации на вѣстникъ-тъ, правителство-то съвсѣмъ го унищожи и запрѣти на Славейкова да издава вече какъвъ-да-е вѣстникъ, а и него самаго тури въ тъмница, по обвинение, че въ послѣдни-тѣ броеве на „Македония“ явно проповѣдвалъ резолюционни идеи между Българе-тѣ.

We used 200 pages from about 1000 book pages scanned at a resolution of 200 DPI. The images are about 2300×3600 pixels (8.28 MPixels), 14.8 x 23.3 cm, grayscale 256 (8 BitsPerPixel). We use preprocessing to convert the images to 1 bit per pixel, black and white, by the help of Image Magic software with 60% threshold value.

The goal of our experiments is to compare practically the efficiency of described methods counting the number of correctly retrieved words in a sequence of words, sorted by their similarity measures with respect to the corresponding HD. For all experiments the same segmentation is used. We choose a pattern word and then measure similarities between it and the words with approximately same width.



Tables contains numbers of correct words in an ordered sequence with the corresponding distance D . m and n in the ratio m/n denote:

- m , the number of correct words with distance D ;
- n , the number of all words with distance D .

For word **ВСИЧКИ**

$D =$ Method	4	5	6	7	8
HD	16/16	44/44	115/120	168/217	177/500
PHD+3	77/77	206/254	209/500	–	–
CHD	19/19	213/252	214/500	–	–

For word **Русия**

$D =$ Method	4	5	6	7
HD+1	2/2	3/3	5/5	5/6
PHD+3	3/3	11/15	–	–
CHD		8/8	13/24	–

We count the number of correctly retrieved words among first 100, 200, ..., 500 words with approximately same width. m is the number of correctly retrieved words among first n words in the ordered sequence in the notation m/n .

For word **всички**

$n =$ Method	100	200	300	400	500
HD01	97	158	186	195	206
MHD	100	169	199	207	212
SHD	100	177	205	213	220
M-HD	100	173	202	214	218
LTS-HD	100	185	215	221	224
NEW	97	164	198	213	224

For word **Русия**

Method			
HD01	4/4	9/18	10/23
MHD	10/10	14/23	15/49
SHD	11/11	14/24	–
M-HD	7/7	12/14	–
LTS-HD	10/10	14/23	–
NEW	7/7	12/15	14/26

There are two relative words (derivatives) of the pattern word **всички**, namely **всичка** and **всичко**. We count as correct words all three of them. This is very useful in practice and show another advantage of methods under discussion and our approach in search. Also, there are 5 similar words of the word **Русия**: **Руски**, **Руска**, **Руско**, **руски** and **руска**.

The best results are in bold in all tables.

Discussion and Conclusion

The main conclusions that we derive from are:

1. “Sum-distances” outmatch “point-distances”.
2. There are no significant differences between the methods that we call “sum-distances” ones.

In this article we do not discuss the quality of image preprocessing particularly the important step of segmentation.

We have no data of number of searching words in the text, because this is tedious work which cannot be done by computer. It follows than we cannot produce the standard recall/precision retrieval estimation.

We think that our comparison of similarity methods is significant for their implementations in software searching systems. In spite of low efficiency of these Hausdorff type methods (the searching takes a lot of time) high level personal computers could be able to solve the problem in reasonable time.

Appendix

Bulgarian typewritten document (about 1940), 335 pages, tif (2400×3200), 1 BPP

I РАЗДЕЛ

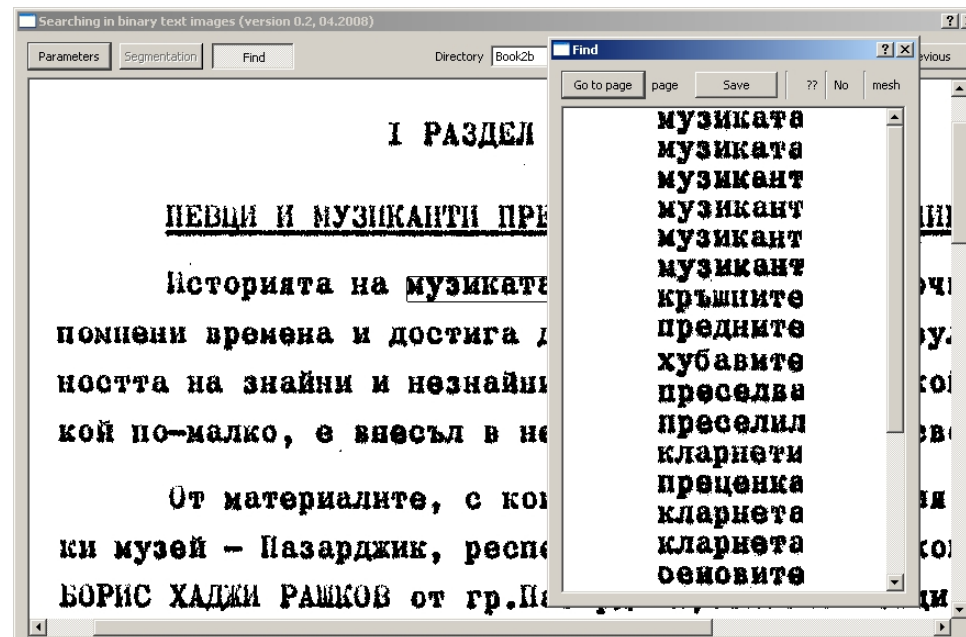
ПЕВЦИ И МУЗИКАНТИ ПРЕДИ И СЛЕД ОСВОБОЖДЕНИЕТО

Историята на музиката в гр.Пазарджик започва от незапомнени времена и достига до наши дни, като резултат от дейността на знайни и незнайни труженици, които, кой повече, кой по-малко, е внесъл в нейната съкровищница своя дял.

От материалите, с които разполага Окръжния исторически музей – Пазарджик, респективно сведенията, които е събрал БОРИС ХАДЖИ РАШКОВ от гр.Пазарджик, относно певци и музиканти преди и след Освобождението се установява, че битовите нужди, свързани с годижи, сватби, занаятчийско-еснафски сбирки, хора, вечеринки и пр. са били задоволявани от музиканти – професионалисти и любители.

Професионалисти били онези музиканти-инструменталисти или певци, като най-често инструменталиста е бил и певец, които са свирили и пеели срещу възнаграждение, а любители – онези, които със своето пееие и свирене са радвали душите и сърцата на хората по сборове, хорища и др., без да получават възнаграждение.

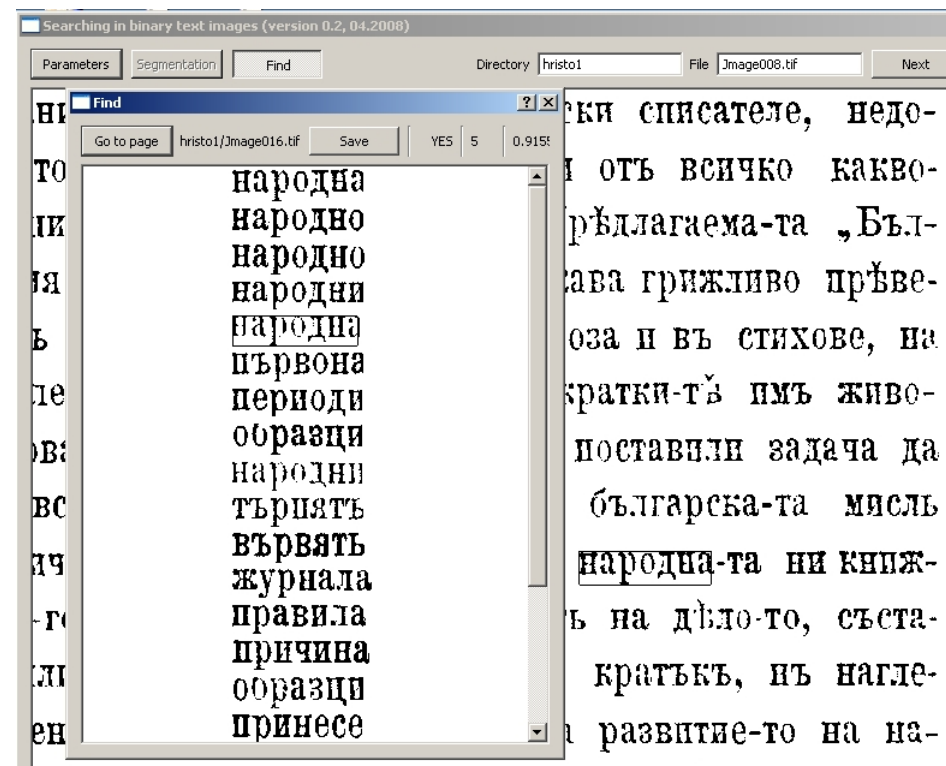
Но и едните и другите със своето майсторство са пренасяли душите на своите слушатели в друг мир. С живите хоровадни мелодии те са карали неспринудено хората да играят хора и ръченици, карали са със своите гласови възможности да забравят делничните трудности, като същевременно са поддържали будно националното съзнание посредством българските напеви и словесно съдържание.



Bulgarian book, Christomatia (1884), 1000 pages, tif (2300×3800), 8 BPP

Повече-то отъ ранни-тѣ му стихотворения сѣ любовни пѣсни, по подражание на грѣцки-тѣ, и не прѣдставляватъ литературна стойностъ; стихотворения-та му въ „Смѣсна Китка“ при всичко, че повечето сѣ слаби подражания на руски-тѣ, нѣ свидѣлствуватъ вече за поетическо-то дарование на г. Славейкова; най-добри-тѣ му стихотворения сѣ обнародвани-тѣ по-послѣ въ „Читалниште“, отъ които „Не пѣй ми се,“ „Жестокостъ-та ми се сломи“ и „Тогасъ повѣ“ джхатъ сѣ истински лиризмъ и заслужено привлѣкоха внимание-то на читателе-тѣ. Славейковъ, който е вѣштъ въ български езикъ, прѣвъ доказа глѣвкостъ-та му въ поезия-та. Като се числи между първи-тѣ борци по черковни въпросъ, той захваша въ сѣмшто-то врѣме почтенно мѣсто въ редъ-тѣ на малко-то ни добри литератори.

Велико влияние е упражнила възъ пробуждане-то духъ-тѣ камъ свобода-та на независимостъ-та у български народъ доста обширна-та литературна дѣятелностъ на *Георгий Сава Раковски* (род. въ Котелъ 1818, умр. въ Букурещъ 1868 г.). Въ личностъ-та и въ дѣла-та на Раковски се отрази най-нагледно тогавашно-то състояние на умове-тѣ, нужди-тѣ, стремления-та и идеали-тѣ на народъ-тѣ ни. Тако-речи единъкъ дѣецъ по онова врѣме, той писува, работи всичко. Той искаше да обгърне въ своя-та широка дѣятелностъ всички-тѣ нужди на народъ-тѣ ни, да удовлетвори всички-тѣ национални купнѣяния, да осмѣстива най-сѣмжни-тѣ и възделени мечти. Той възсъздаде сѣ фанатически въсторгъ минжло-то и приготви бжджште-то. Бѣше въ сѣмшто-то врѣме поетъ, историкъ, етнографъ, публицистъ, агитаторъ и хайдутинъ. Нито на единъ български дѣятелъ животъ-тѣ не е билъ напълненъ сѣ толкова неутолима и разнообразна дѣятелностъ и напѣстренъ сѣ толкова бѣди, приключения и странности. Той се бѣше училъ въ Атина, Парижъ, Цариградъ и въ Русия. Знаеше руски, сръбски, румжнски, турски, грѣцки, староелински, французскѣй, арабски и дори отъ чѣсти санскрит-

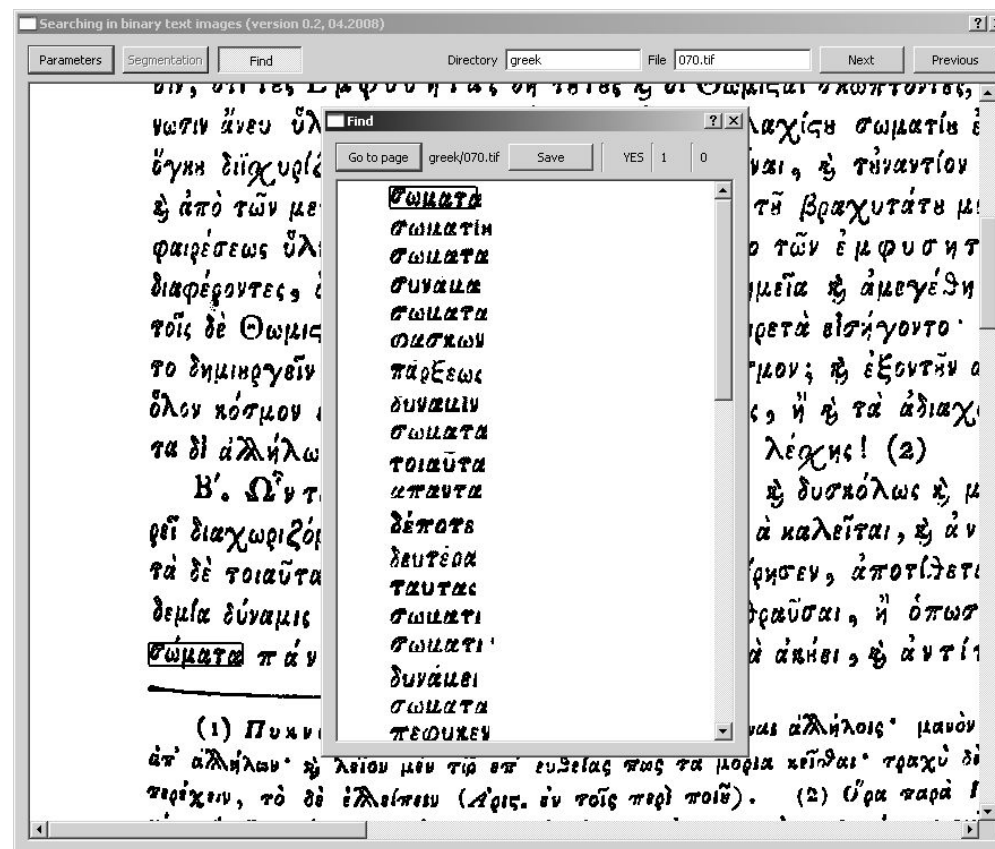


Old Greek text (approx. in the third century BC), 50 pages, jpg (1077×1416), 8 BPP

τοιαῦτα κατὰ σχῆμα πάντα εἰσὶν ἄτρεπτα, οἷα μ
Σῶμα δὲ σύνθετον ἐν τῇ φύσει ἔδεν τοιοῦτον· ἥπ
θλίψει, τομῇ ὅπως ὑπείκει, καὶ λίθους οἱ σερρ
τητι διαφέρων ἀδάμας, μηδενὸς ὅλως ἐξαιρημένον,
μοιρον, ὥς δέδεικται. Ὡς δὲ σωμάτων τὰ μέρη
καὶ ῥᾶσα διαχωρίζεσθαι πέφυκεν ἀπαλὰ ταῦτ
μέλι, ἄργιλλος, κτ· ὅπως δὲ ἦττον καὶ ἁτονωτέρω
πλοκῆς ἀποσείχει, τοσῆτι καὶ ἀπαλώτερα, ἐς ἅκρ
ἔδεν γὰρ ἐσὶν ἐντυχεῖν, οὗ τὰ μέρη μὴ ὅπως ὄν
γνύμενα.

Γ'. Τὸ σκληρὸν σῶμα ὑπὸ κήφης τε καὶ ἀοιδ
πεφυκός, εὐθραυστον ἀκίει· τοιαῦτα χάλυψ
τὰ κεράμεια σκεύη· Τέτων τὰ σερρεὰ μέρη ἔχ
ἐλήλοις, διό καὶ ῥᾶσα τῆς ἀμοιβαίας παφῆς ἀφίστα
σώματα πάν

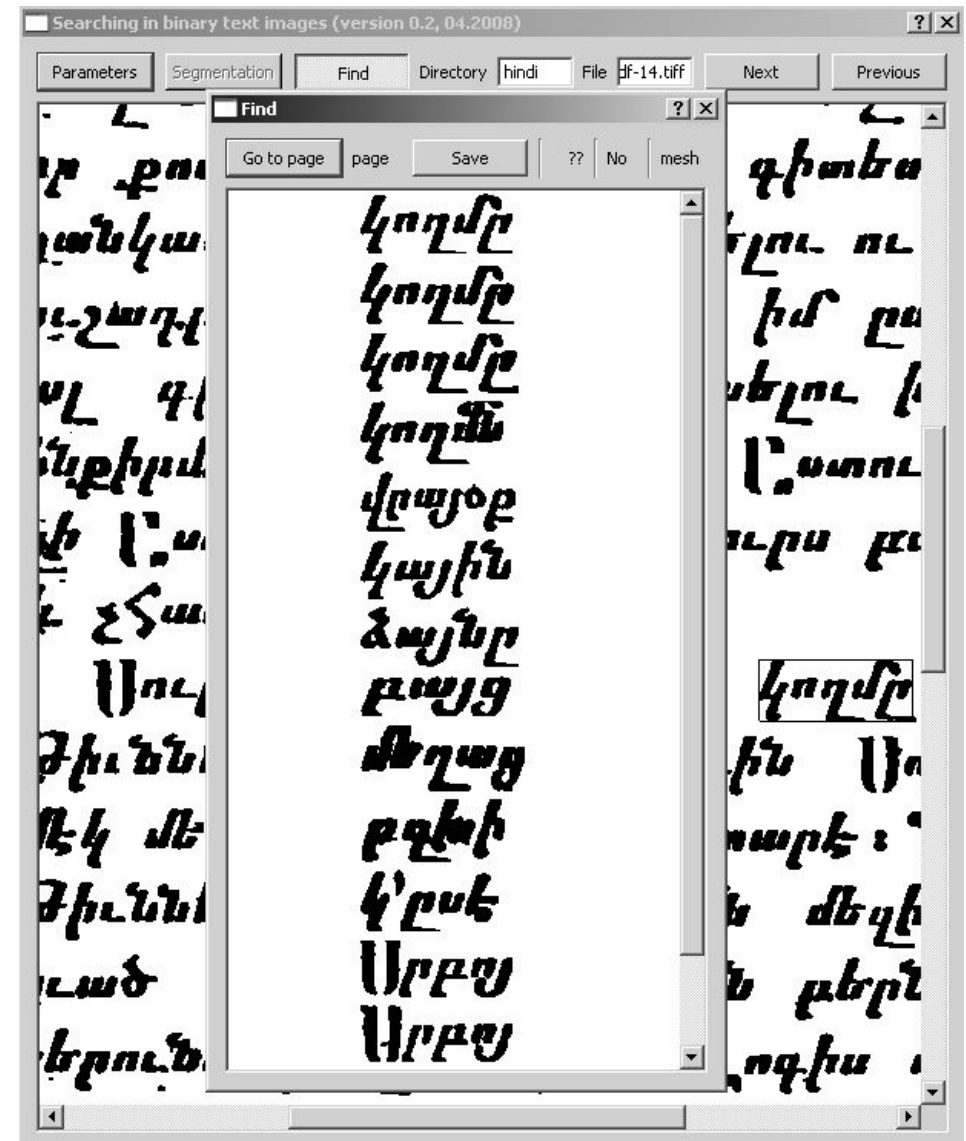
Δ'. Τὸ ἐκ πολλῶν οἶονεῖ λεπίδων πάνυ λεπτῶν ἀ
εἰσόμενον, εὐσχιστον σῶμα προσείρηται· τέτοι
μόρια σερρότερον προσκεκόλληται ἀλλήλοις, ἢ λεπίς
ἄρα τὰ τοιαῦτα σώματα εἰς λεπίδας ἀναλυόμενα·
καλόμενοι λίθοι οἱ ἐξ Ἰβηρίας, καὶ Καππαδοκίας,



Armenian book (1858), 178 pages, tiff (2800×5000), 1 BPP

կառավարութեան ոգին ամէն աստիճանի մարդոց սրտին մէջ տարածուեցաւ, և սորվեցուց անոնց մէկզմէկ ասելու ու մէկզմէկէ գարշիլ. մինչև անգամ մէկ խանութի մէջ գործող արհեստաւորը սորվեցաւ նզովք տալ իր քովի խանութին մէջ բանող դրացիին, ան պատճառով որ անիկայ Հոգւոյն Սրբոյ բղխումը իրեն համաձայն չդաւանիր. ոչ մէկը և ոչ մէկալը հասկընալով թէ ինչ կ'ըսեն, կամ ինչ բան կ'ուզեն հաստատել:

Ուստի այսպիսի անտեղի վիճաբանութիւնները պատճառ տուին որ երբոր մարդիկ Հոգւոյն Սրբոյ վրայօք խորհին, գրեթէ միայն աս մէկ նիւթիս ուղղեն միտքերնին, այսինքն թէ՛ Հոգւոյն բղխումը միայն Հօրմէն է, կամ Հօրմէն ու Որդիէն: Եւ մէն ալ կը դաւանին թէ Հոգին Սուրբ՝ Երբորգութեան մէկ անձն է. բայց ո՞վ կրնայ ըսել թէ անիկայ ինչ ներգործութիւն կ'ընէ մարդու հոգւոյն փրկութեանը համար, կամ ինչ է իր մասնաւոր պաշտօնը մարդս երկինքը բարձրացընելու համար: Եւս աս մեծ և ամենահարկաւոր նպատակիս համար է որ Երրորդութեան վարդապետութիւնը յայտնուած է: Հայրը խրկեց Որդին աշխարհը փրկելու: Վնչու որ Եւսուած անանկ սիրեց աշխարհը մինչև որ իր միածին Սուրբ



Text in Spanish (1901), 30 + 57 pages, gif (1400×2500), 4 BPP

Alonso, Rogelio M., Cartilla histyrico-descriptiva del tĩrmino municipal de Macuriges. Habana: Impr. La Propagandista, 1901, HOLLIS Catalog, Harvard University

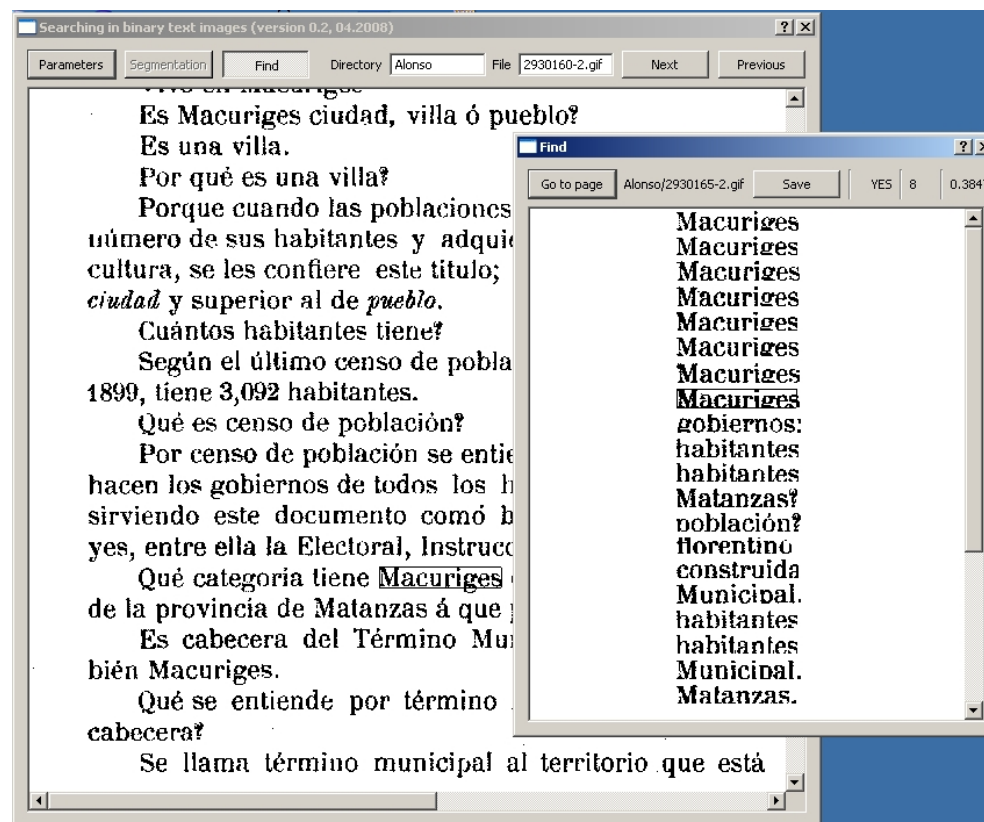
No señor; todas las fincas azucareras tienen sus chuchos que conectan con las líneas del ferrocarril y hay adcmás caminos reales, trasversales y vecinales, estos en estado natural. (1)

Qué entiende V. por caminos reales, trasversales y vecinales?

Caminos reales, son los caminos abiertos por el gobierno Español desde los tiempos primeros de la colonización de la Isla de Cuba y tienen de ancho 24 varas; caminos trasversales son los que solo tienen de ancho 12 varas y vecinales los pasos permitidos por los propietarios de fincas, para acortar distancias de un lugar á otro y salvar lo mal que pudieran estar los caminos por el fango, las piedras ó la yerba.

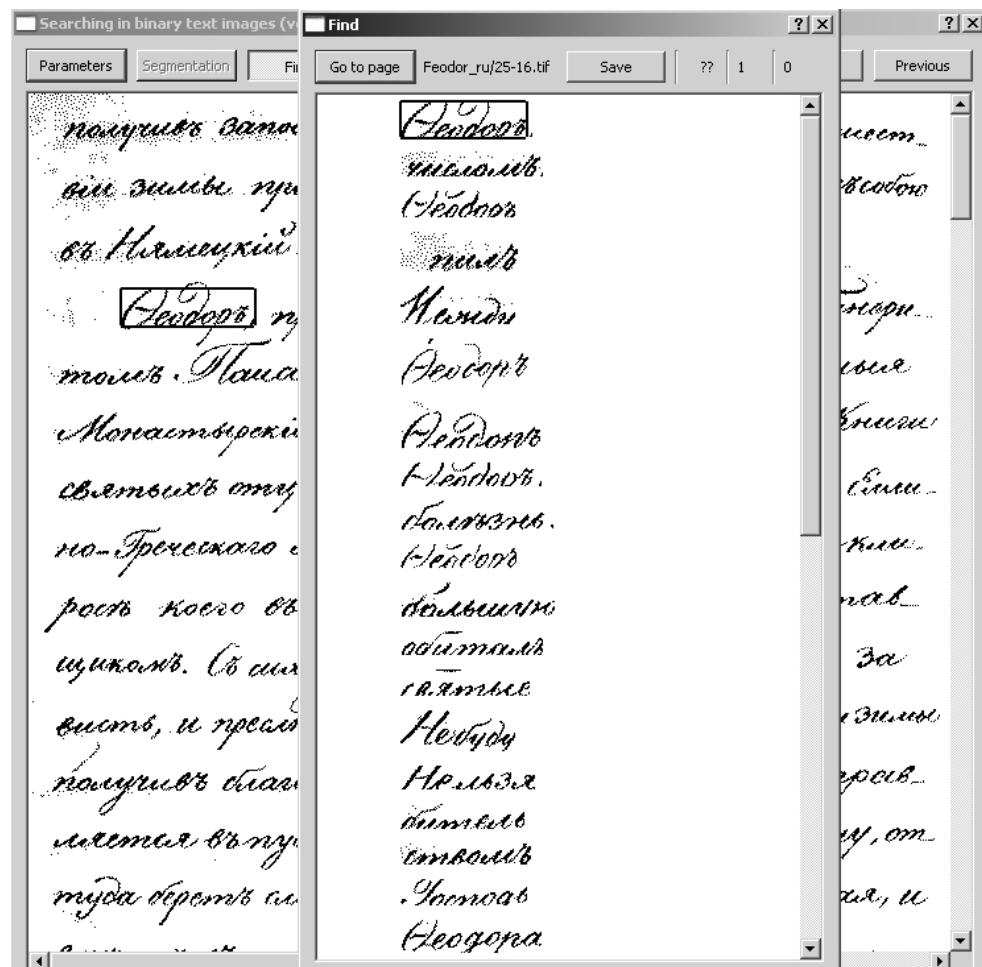
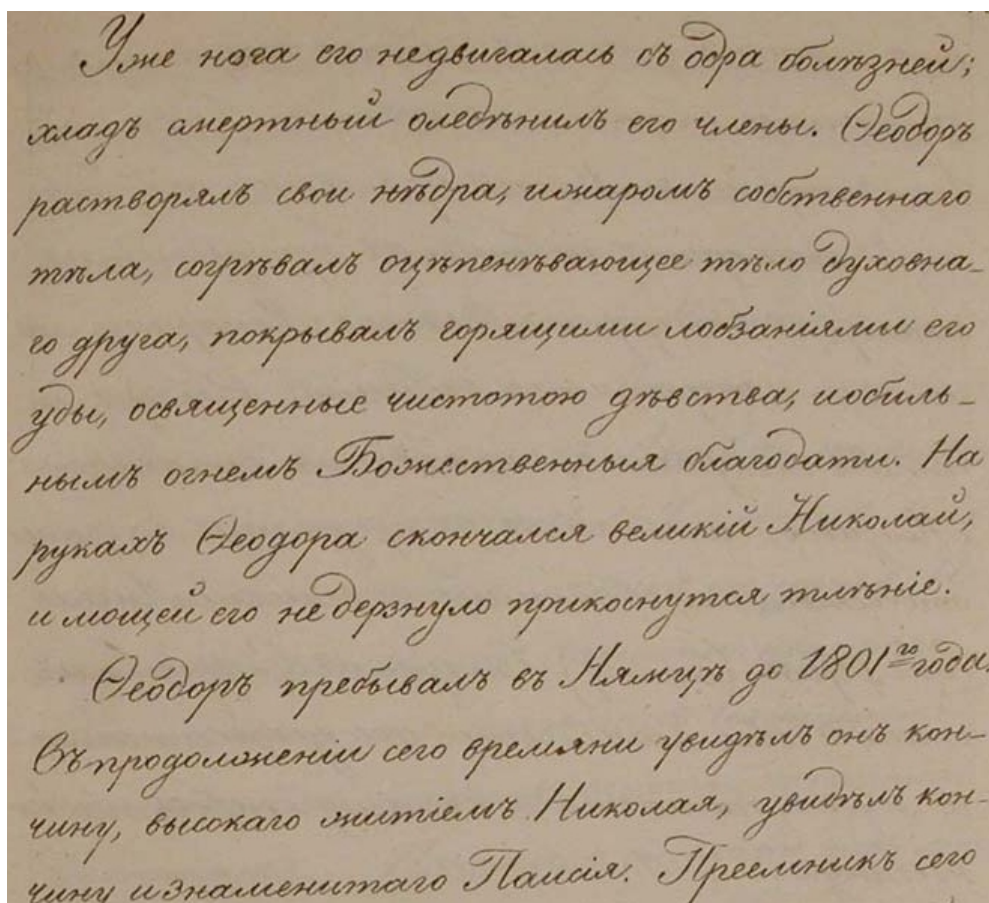
Cuántos ingenios para la fabricación de azúcar tiene en la actualidad el Término todo?

Los siguientes: «Santa Filomena» en el barrio de Navajas propiedad del Sr. Leandro Soler, «Elizalde» del Sr. Alberto Broch en el Ciego y «Santa Catalina» del señor Enrique Heedigg en el mismo barrio; «Carmen» del Sr. Alexander en Navajas, «Socorro» del Sr. Pedro Arenal en Tramojos y «Dolores» del Sr. Francisco Rosell en Platanal, todos centrales y con magníficos aparatos.



Handwritten document in Russian (1840), 44 pages, jpg (700×900), 24 BPP

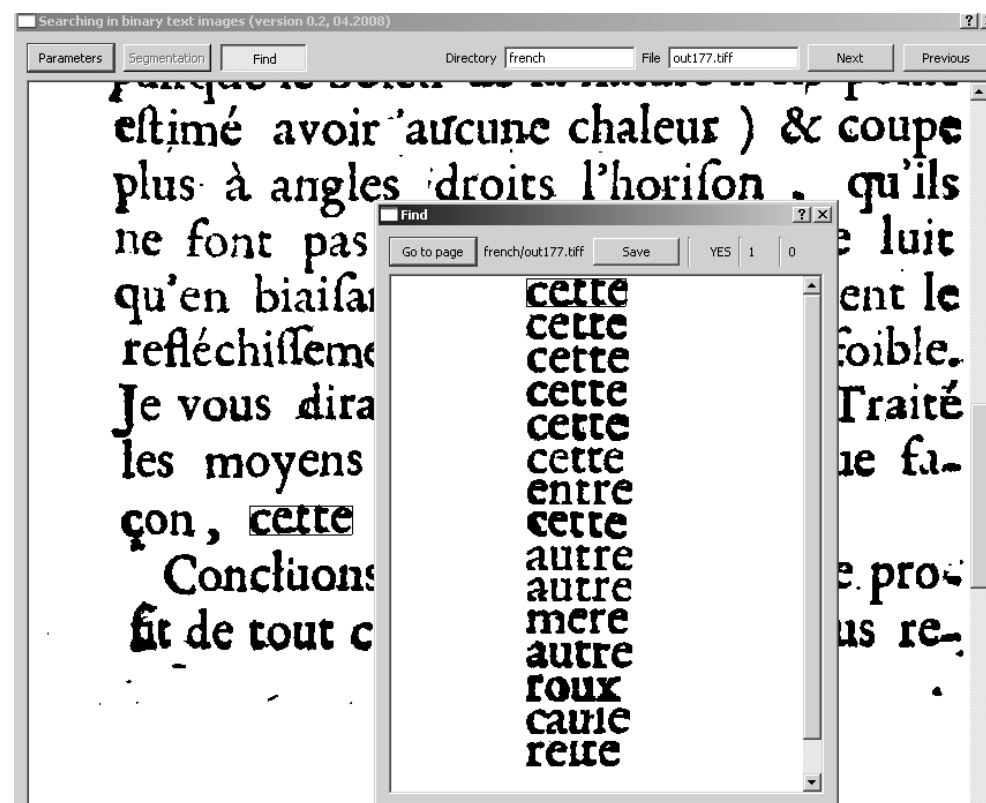
Дом живоначальной Троицы, Свято-Троицкая Сергиева Лавра, Собрание славянских
рукописей, 43: Житие схимонаха Феодора



Text in French (1692), 388 pages, jpg (2048×3550), 8 BPP

Nicolas de Bonnefons, Ch. de Sergy, (1692), University of Gent, Digitized by Google (2007)

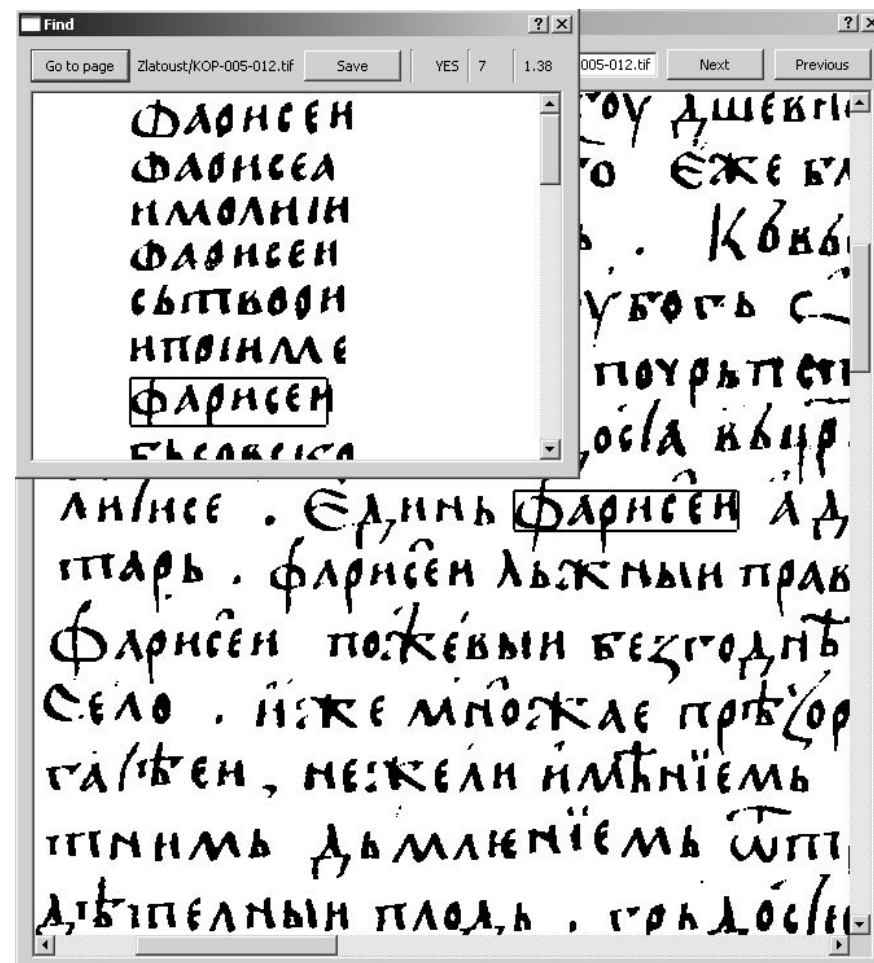
Quant à la Terre, si vous la rencontrez
bonne, ce vous fera un grand avantage, &
une grande épargne ; mais rarement en
pourrez-vous trouver, où il n'y ait beau-
coup à travailler, d'autant que telle pa-
roîtra passablement bonne au dessus, qui
étant ouverte de la profondeur d'un fer
de Béche seulement, se trouvera Argi-
leuse dessous ; ce fonds est pire aux Ar-
bres que le Tuf, ou la Roche, à cause qu'il



Slavonic manuscript, (1574), 747 pages, jpg (1249×1890), 24 BPP.

Дигитална Народна библиотека Србије, Њирилски рукописи, Збирка словенских рукописи Јернеја Копитара, Зборник “Златоуст”

сѣмрътнѣ имѣхѣ . ꙗко да не ма дѣюще се
боудѣмъ насѣ , ꙗ ꙗба въскръшающаго
мрътвые . ꙗже ѿполныи сѣмрътнѣ ии
збави ꙗ ии збави . ꙗ ꙗже оуповахѣмъ
ꙗко иеще ии збави . Упо оубо рци ми
възносишѣ ꙗко ѿ добрыхъ своихъ . Сѣмъ
спожѣ блгодѣти исповѣдовахѣ да хощомъ
упо бо имашѣ еже непріель еси . ꙗже
ли пріель еси , упо се хвалиши ꙗко непріе
мъ . не ты ба познахѣ еси правдою , ꙗ
въ тебѣ блгостію позна . ꙗже рече ба
паче же познахѣ бывше ѿ ба . не ты ба
пріель еси добродѣтелию . ꙗ тебѣ хс
пришьствіемъ пріепи . гонѣ бо рече ꙗже
и поспигнухѣ имъ же и поспигахѣ быхѣ
ѿ хс . не вы мене ии збрасахѣ рече . ꙗ
азъ ии збравасѣ . ꙗ ли зане похъпѣхѣ еси
веле мудрствовашѣ ; имашѣ въвину



Thank you for your attention.