Impossible & Advanced Optical Character Recognition Using a Cheap Point & Shoot Digital Camera

INTRODUCTION

This study was conducted using a point & shoot digital camera *Konica Minolta DiMAGE X1* with 8.1 Megapixels with no previous adjustment at 3264x2448 pixels with 72 dpi. Samples from 4 books were used:

1. [MILTON, JOHN.] The Grand Case of Conscience concerning the Engagement Stated & Resolved.

printed in London 1650

2. [PETKOV-MISIRKOV, KRSTE.] Za makedonskite raboti.,

printed in Sofia 1903

3. [DE MONTESQUIEU, MONSIEUR.] Oeuvres. Tome I,

printed in London 1787

4. [VARIOUS AUTHORS] Macedonian Review, year IV, book#4,

printed in Sofia 1933

Only JONH MILTON book was "scanned" at Alexander Library at Rutgers, the State University of New Jersey with expensive professional camera scanner. It is used here for the sake of the comparison in relation to the curve of the pages and the **high resolution professional cameras Vs cheap point & shoot digital cameras.**

The "scanning" was performed using a light bulb of 50W in room conditions. Some pattern training was done using Abbyy Finereader 6.0 Different compression methods were used to reduce the file size to its minimum web centricity. Optical Character Recognition engines require the minimum of 150 dpi (the optimum is 300 dpi) for accurate results. Using an experimental program with Abbyy Finereader 5.0 engine at 72 dpi JPEG (instead of TIF) files some amazing results appeared!

RESULTS

When making a full text search of the files by using the actual alphabet, the recognized text is given/written in odd or non-existing characters. Having in mind that very cheap equipment was used in everyday (ordinary) conditions, the quality of the extracted text is excellent. There is no explanation needed in this case (please look at the screenshots on the poster very carefully.)

CONCLUSION

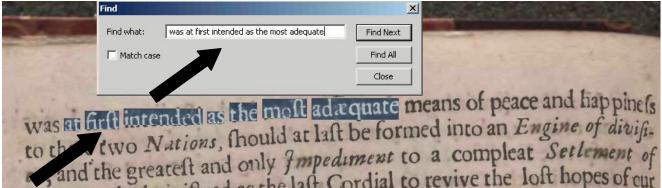
It is possible to extract the text by using a cheap point and shoot digital camera with 8 MP and by some previous experimentation of the light and the curvature of the pages. If you consult an Optical Character Recognition Specialist, it will save you a lot of trouble and time when you scan or photograph rare and old books. Most of the libraries, national archives and museums use digital cameras instead of scanners to "scan" the books. It is very complicated to extract the text out of them, because of the 72dpi resolution and the "no adjustments features" made by the person who uses the expensive equipment for photographing the book.

FUTURE

There is more to be done in using the digital camera to digitalize a book. The high resolution camera scanner does not guarantee the best results. We need more research in the field of the scanning and the right resolution, curve of the pages, ISO settings, exposure time and bigger lenses.

Can 72 dpi (half than the minimum for OCR) of the point & shoot be better than expensive and sometimes slow scanners? **Yes!** Do we need special conditions to increase the quality of the OCR? **No, we don't!** Library users, in a year or two, will be able to digitize a book at their library work desk using a cheap digital camera, free software or free internet conversion servers to DJVu format or demo version of the ABBYY Finereader software (ScanR servers uses only 2 MP pictures from the mobile phone for OCR).

Also, please remember: OCR done with the digital camera is at 72dpi...The minimum is 150 dpi, optimum 300dpi. The future of OCR is right here-right now...Thank you!



Peace, and administred as the last Cordial to revive the lost hopes of our Peace, and administred as the last Cordial to revive the lost hopes of our dying Enemies? And after we have (through God) struck out all the dying Enemies? And after we have, (through God) struck out all the Carnal weapons out of their hands, yea, separated the Head from that Sample from [MILTON, JOHN.] The Grand Case of Conscience concerning the

Sample from [MILTON, JOHN.] The Grand Case of Conscience concerning the Engagement Stated & Resolved. printed in London 1650.

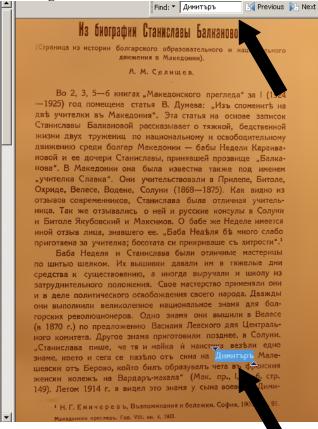
Professional scanner camera is used. Some of the characters are not well recognizes by the OCR machine. Still, the book can be searched in modern English characters.

Screen 1 of 2

Из биографии Станиславы Балкановой. (Страница из истории болгарского образовательного и национального движения в Македонии).

Я. М. С.е лищев.

Во 2, 3, 5-6 книгах "Македонского прегледа" за I (1924 – 1925) год помещена статья В. Думева: "Изъ споменитъ на двъ- учителки въ Македония". Эта статья на основе записок Станиславы Балкановой рассказывает о тяжкой, бедственной жизни двух тружениц по национальному и освободительному движению среди болгар Македонии — бабы Недели Караивановой и ее дочери Станиславы, принявшей прозвище "Балканова". В Македонии она была известна также под именем учителка Славка". Они учительствовали в Прилепе, Битоле, Охриде, Велесе, Водене, Солуни (1868—1875). Как видно из отзывов современников, Станислава была отличная учитель ница. Так же отзывались о ней и русские консулы в Солуни и Битоле Якубовский и Максимов. О бабе же Неделе имеется иной отзыв лица, знавшего ее. "Баба Неделя бЪ много слабо приготвена за учителка; босотата си прикриваше съ хитрости". 1 Баба Неделя и Станислава были отличные мастерицы по шитью шелком. Их вышивки давали им в тяжелые дни средства к существованию, а иногда выручали и школу из затруднительного положения. Свое мастерство применяли они и в деле политического освобождения своего народа. Дважды они выполнили великолепное национальное знамя для болгорских революционеров. Одно знамя они вышили в Велесе (в 1870 г.) по предложению Василия Левского для Централь ного комитета. Другое знамя приготовили позднее, в Солуни. "Станислава пише, че тя и майка и наистина везвли едно знаме, което и сега се паз1зло отъ сина на Димитъръ Малешевски отъ Берово, който билъ образувалъ чета въ френския женски колежъ на Вардаръ-махала" (Мак. пр., І, 5-6, стр. 149). Летом 1914 г. я видел это знамя у сына воеводы Дими-1 Н. Г. Еничеревь, Възпоминания и бележол. София, 1906, стр. 91. Македонски прегледь. Год. VIII, кн. 4. 1933.



Sample from [VARIOUS AUTHORS] Macedonian Review, year IV, book#4, printed in Sofia 1933. Made with point & shoot digital camera *Konica Minolta DiMAGE X1* with 8.1 Megapixels with no previous adjustment at 3264x2448 pixels with 72 dpi.

For the sake of the comparison on the left is extracted text and on the right the full text search original.

Advanced Optical Character recognition done with point and shoot digital camera at 72dpi. No explanation needed – just look very carefully at the screenshots samples.

No explanation needed – just look very carefully at the screenshots s
даваат да се сојединиме. Од македонико глединиче соје-
динеін'ето на целя Македонија со Бугарија, или Србија,
или Грција не јет желател'но, но не јет и страшно.
Значит нема зашчо да се бориме со него. Но такво ссіе-
динејн'е не ке допушчат, ни малите балканцки, ни голе-
мите европеіцки држан. Значит, ниіе, не сакаіки да
бркаме нашите интереси со бугарцките, дааме својето
согласіе и Find . Се пи-
ТАТ САМО, Find: - Бугарија 🔄 Previous 🛌 Next 110.138, 38-
шчо кажува у арцотани,
и ке ни напраела ушче. Да видиме, шчо арно сме ви-
деле није до сега од бугарите?
Бугарите усилиіа біуджегот на екзархината от по-
іавуаін'ето на србцката пропаганда; со друзи зборон
усилија нивната пропаганда и нивните интереси во Ма-
даваат да се соједнытие. Ол македонцко гледицие соје-
динеін'ето на цела Македонија со Бугарија, или Србија,
нли Грција не јет желател'но, по не јет и страшно.
THE PULL POTE
мите евр Find: * македонија 🔄 Previous 🎦 Next Сакајки да
бркаме нашите интереси со бугарцките, дааме својето
согласіе в узаконуаме сашчесчествујушчијо ред. Се пи-
тат само, аля тоа узаконуаін'е іст во наша полза, за-
шчо кажуваат, оти Бугарвія ни напраила многу арцотаін,
и ке ни напраела ушче. Да видиме, шчо арно сме ви-
даваат да се сојединиме. Од македонцко гледишче соје-
динејк сто на цела Македонија со Бугарија, или Србија,
нли Грција не јет желател'но, но не јет и страшно.
Значит нема зашчо да се бориме со него. Но такво ссіе-
динејн'е не ке допушчат, ни малите балканцки, ни голе-
мите европеіцки држан. Значит, ниіе, не сакаіки да
бркаме нашите интереси со бугарцките, дааме својето
согласіе и Find
ТАТ САМО, Find: - Срција Previous Next ПОЛЗА, За-
тат Само, Find: - Срција Mext Next нолза, за-
и ке ни напраела ушче. Да видиме, шчо арно сме ви-
деле није до сега од бугарите?
[PETKOV-MISIRKOV_KRSTE] Za makedonskite raboti _ printed in Sofia 1903

[PETKOV-MISIRKOV, KRSTE.] Za makedonskite raboti., printed in Sofia 1903

Advanced Optical Character recognition done with point and shoot digital camera at 72dpi. No explanation needed – just look very carefully at the screenshots samples.

 $T_{\rm ELS}$ font les principes des trois gouvernemens : ce qui ne fignifie pas que, dans une certaine république, on foit vertueux; mais qu'on devroit l'être. Cela ne prouve pas non plus que, dans une certaine monarchie, on ait de l'honneur : & que, dans un état despotique particulier, on ait de la crainte; mais qu'il faudroit en avoir : fans quoi le gouvernement fera imparfait.

rinu		
Find: 🔻	sans	🔣 Previous 📡 Next

 $T_{\rm ELS}$ font les principes des trois gouvernemens : ce qui ne fignifie pas que, dans une certaine république, on foit vertueux; mais qu'on devroit l'être. Cela ne prouve pas non plus que, dans une certaine monarchie, on ait de l'honneur : & que, dans un état defpotique particulier, on ait de la crainte; mais qu'il faudroit en avoir : fans quoi le gouvernement fera imparfait.

Find Toolbar			
Find: 🔻	mparfait	🔄 🔤 Previous 📐 Next	

LELS font les principes des trois gouvernemens : ce qui ne fignifie pas que, dans une certaine république, on foit vertueux; mais qu'on devroit l'être. Cela ne prouve pas non plus que, dans une certaine monarchie, on ait de l'honneur : & que, dans un état defpotique particulier, on ait de la crainte; mais qu'il faudroit en avoir : fans quoi le gouvernement fera imparfait.

Find: • despotique

[DE MONTESQUIEU, MONSIEUR.] Oeuvres. Tome I, printed in London 1787

🖂 Previous 🔀 Next

Advanced Optical Character recognition done with point and shoot digital camera at 72dpi. No explanation needed – just look very carefully at the screenshot

Изъ културното и обществено минало на българитъ въ Банско. Оть Костадинь Г. Молеровь.

Тамъ, гдето слънцето на залѣзъ озарява съ кървава свѣтлина отвеснитѣ канари на вѣчноснѣжния "Сухдолъ" и бърза да се закрие задъ тѣхъ като задъ приказни порти; гдето неговитѣ лжчи въ предсмъртна агония скачатъ отъ "Ку́тела" на "Елъ-тепе" и опиратъ чакъ до отсрещнитѣ зжбери на острия "Газѣй" и на дивния "Кралевъ дворъ", тия и много други великани на Пирина редятъ своитѣ снѣжни глави като бисери на венецъ, а подъ тѣхъ сини езера отразяватъ дивностъта на великата планина и проточени въ рѣка "Глазне" шумятъ и пѣятъ ведно съ гори и долове мила за душата пѣсень — тамъ е днешно Банско. Свидна рожба на Пирина, родени съ него, така да се каже, единъ за другъ, Банско е пазитель и пѣвецъ на неговата легенда.

Но пръснат	Find			× ×	
лина, к	c Find: 🔻	5ългарите	Previous 🕅	Next	жката до- 1 обширно
			т в на пирина, Рі і прошарено тукт		
селища	и махал	и. Троялищат	а ¹ около Банско:	CB.	Илия, Св.

Изъ културното и обществено минало на българитъ въ Банско.

Оть Костадинь Г. Молеровъ.

Тамъ, гдето слънцето на залѣзъ озарява съ кървава свѣтлина отвеснитѣ канари на вѣчноснѣжния "Сухдолъ" и бърза да се закрие задъ тѣхъ като задъ приказни порти; гдето неговитѣ лжчи въ предсмъртна агония скачатъ отъ "Кутела" на "Елъ-тепе" и опиратъ чакъ до отсрещнитѣ зжбери на острия "Газѣй" и на дивния "Кралевъ дворъ", тия и много други великани на Пирина редятъ своитѣ снѣжни глави като бисери на венецъ, а подъ тѣхъ сини езера отразяватъ дивностъта на езликата планина и проточени въ рѣка "Глазне" шумятъ и пѣятъ ведно съ гори и долове мила за душата пѣсень — тамъ е днешно Банско. Свидна рожба на Пирина, родени съ него, така да се каже, единъ за другъ, Банско е пазитель и пѣвецъ на неговата легенда.

Но п	реди да	заеме днец	июто си мъсто,		
пръснато і	Find				ката до-
лина, коят платно, за	Find: 🔻	јеятъ	Previous	Next	бширно
покрито с	S FOON H	Dacoutta u	TOOLUS DOULO THE		

Изъ културното и обществено минало на българитъ въ Банско.

Оть Костадинь Г. Молеровъ.

Тамъ, гдето слънцето на залѣзъ озарява съ кървава свътлина отвеснитъ канари на вѣчноснѣжния "Сухдолъ" и бърза да се закрие задъ тѣхъ като задъ приказни порти; гдето неговитѣ лжчи въ предсмъртна агония скачатъ отъ "Кутела" на "Елъ-тепе" и опиратъ чакъ до отсрещнитѣ зжбери на острия "Газѣй" и на дизния "Кралевъ дворъ", тия и много други великани на Пирина редятъ своитѣ снѣжни глави като бисери на венецъ, а подъ тѣхъ сини езера отразяватъ дивностъта на великата планина и проточени въ рѣка "Глазне" шумятъ и пѣятъ ведно съ гори и долове мила за душата пѣсень — тамъ е днешно Банско. Свидна рожба на Пирина, родени съ него, така да се каже, единъ за другъ, Банско е пазитель и пъвецъ на неговата легенда. Но преди да заеме днешното си мѣсто, Банско е било

пръснато г					ката до	
лина, коят платно, за	Find: *	ревецъ	Previous	🔊 Next	бширн допиті	
покрито ст селища и	ь гори и махали,	пасоища и п Троялищата ¹	около Банско	ъ-таме ст	малк	

[VARIOUS AUTHORS] Macedonian Review, year IV, book#4, printed in Sofia 1933