

Persian and Arabic Optical Character Recognition for Open Source: Experiences of the FID Nahost, Halle (Saale)

45th Annual Conference of MELCom International

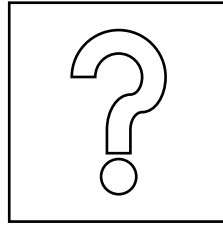
Maida Farooq Buttar

London, 16.05.2024



OCR Engine

Kraken



Tesseract

- Open source
- Suitable for RTL languages
- Suitability for unseparated letters such as Arabic
- Existing support for the languages we need (Arabic, Persian, etc.)
- Existing adaptation possible (adaptation from training)
- Larger development community, community support
- Usable results

OCR Tools

Transkribus:

The screenshot displays the Transkribus OCR interface. On the left, there is a 'Thumbnail Overview' showing a grid of document pages, with page 3 highlighted. The main window shows a document with Persian text. Several words and phrases are circled in red, and arrows point to them from a list of corrections at the bottom. The text in the main window is: 'بگذرد تازگی ند ارد. شما بعنوان نماینده قانون بنام ملت آزادی، رابه پشت میز محاکمه جلب کرده اید که بنابه تشخیص آقای بیگانه پرستی، داشتن افکار مضرو خطرناک حتی باتهام سو، زهی عدالت. آقای دادستان بنام قانون و بنام عدالت برای حکم اعدام نموده اند. باور کنید همانقدر که شما چه در موقع درخواست ام و بی تفاوت هستید من نیز بهمان اندازه آرام و بی اعتنا هستم. مقصر بدانم. زندگی برای من مدتهاست که دیگر دور من و آنچه وجود من، خواسته های من، آرمانها و امیدهای من، دلهره ها، ماتم، خواسته های ملت، آرمانها و امیدها، دلهره ها و دلخوشی. اگر زندگی بایستی عبارت از خوردن چند لاش اضافی گوشت و نبودنش هزار بار از بودنش بهتر است. محرک من در بیان آخرین'.

4-4 دادستان و سازمان امنیت بجا سوسی، بیگانه پرستی، داشتن افکار مضرو خطرناک حتی باتهام سو
4-5 قصد متهمند. زهی حقیقت، زهی آزادی، زهی عدالت. آقای دادستان بنام قانون و بنام عدالت برای
4-6 من و دوستم از دادگاه تقاضای صدور حکم اعدام نموده اند. باور کنید همانقدر که شما چه در موقع درخواست
4-7 اینگونه احکام و چه در موقع صدور آنها آرام و بی تفاوت هستید من نیز بهمان اندازه آرام و بی اعتنا هستم.
4-8 آنگاه دل من میتوانست بلرزد که خود را مقصر بدانم زندگی برای من مدتهاست که دیگر دور من و آنچه
4-9 تنها مربوط بشخص من است دور نمیزند، وجود من، خواسته های من، آرمانها و امیدهای من، دلهره ها
4-10 و دلخوشی های من همه و همه در وجود ملت، خواسته های ملت، آرمانها و امیدها، دلهره ها و دلخوشی.
4-11

OCR Tools

LAREX:

The screenshot displays the LAREX OCR tool interface. The main window shows a document with Arabic text. The text is segmented into lines, and some words are highlighted in green. The interface includes a top menu with 'SEGMENTS', 'LINES', and 'TEXT' tabs. Below the menu, there are zoom controls for 'Image Zoom' and 'Text Zoom', both set to 100%. On the left side, there is a vertical list of document pages, with the current page highlighted. On the right side, there is a character selection panel with buttons for 'LOAD', 'SAVE', and 'PRESET'. The panel contains a grid of characters, including Latin letters, Greek letters, and special characters. Below the character grid, there are settings options: 'Settings', 'Show Diff' (checked), 'Only show mismatching lines' (unchecked), and 'Show Prediction' (unchecked). At the bottom right, there are buttons for 'SAVE RESULT' and 'LOAD RESULT'.

00000124_ara

Page

Image Zoom 100%

Text Zoom 100%

94

96

الاسير الذي وهبت لك حتى أُخْبِرُهُ وَأَعْطِيَهُ فَقَالَ قَدْ مَضَى فَارَسِلْ الْكَلْبِي
الاسير الذي وهبت لك حتى أُخْبِرُهُ وَأَعْطِيَهُ فَقَالَ قَدْ مَضَى فَارَسِلْ الْكَلْبِي

الاسير الذي وهبت لك حتى أُخْبِرُهُ وَأَعْطِيَهُ فَقَالَ قَدْ مَضَى فَارَسِلْ الْكَلْبِي

في اثره فلم يَلْحَقْهُ وَقَالَ الْاَعْشَى وَهُوَ زَعَمَ ان سَلِيمَانَ بْنِ دَاوُودَ هُوَ الَّذِي
بْنِ اِثْرَهُ فَلَمْ يَلْحَقْهُ وَقَالَ الْاَعْشَى وَهُوَ زَعَمَ ان سَلِيمَانَ بْنِ دَاوُودَ هُوَ الَّذِي

في اثره فلم يَلْحَقْهُ وَقَالَ الْاَعْشَى وَهُوَ زَعَمَ ان سَلِيمَانَ بْنِ دَاوُودَ هُوَ الَّذِي

بَنَى الْاِبْلَاقَ الْفَرْدَ بَعْدَ ان ذَكَرَ الْمُلُوكَ الَّذِيْنَ اَفْتَنَاهُمُ الدَّهْرَ فَقَالَ
بَنَى الْاِبْلَاقَ الْفَرْدَ بَعْدَ ان ذَكَرَ الْمُلُوكَ الَّذِيْنَ اَفْتَنَاهُمُ الدَّهْرَ فَقَالَ

بَنَى الْاِبْلَاقَ الْفَرْدَ بَعْدَ ان ذَكَرَ الْمُلُوكَ الَّذِيْنَ اَفْتَنَاهُمُ الدَّهْرَ فَقَالَ

ولا عَادِيَا لَمْ يَمْتَنِعِ الْمَوْتَ مَالُهُ وَوَزِدُ بَنِيَامِ الْيَهُودِيِّ اِبْلَاقُ
ولا عَادِيَا لَمْ يَمْتَنِعِ الْمَوْتَ مَالُهُ وَوَزِدُ بَنِيَامِ الْيَهُودِيِّ اِبْلَاقُ

ولا عَادِيَا لَمْ يَمْتَنِعِ الْمَوْتَ مَالُهُ وَوَزِدُ بَنِيَامِ الْيَهُودِيِّ اِبْلَاقُ

بناه سَلِيمَانَ بْنِ دَاوُودَ حَقِيْبَةً لَهُ اَرْجُحُ عَلٍ وَتَلْسِي مُسَوِّفُ
بناه سَلِيمَانَ بْنِ دَاوُودَ حَقِيْبَةً لَهُ اَرْجُحُ عَلٍ وَتَلْسِي مُسَوِّفُ

بناه سَلِيمَانَ بْنِ دَاوُودَ حَقِيْبَةً لَهُ اَرْجُحُ عَلٍ وَوُطِي مُوْتَفِقُ

ابوازي كُبَيْدَاتِ السَّمَاهِ وَدَوْنَهُ بِلَاطُ وَدَارَاتِ وَكَلْسُ وَخَنْدُقُ
ابوازي كُبَيْدَاتِ السَّمَاهِ وَدَوْنَهُ بِلَاطُ وَدَارَاتِ وَكَلْسُ وَخَنْدُقُ

LOAD SAVE PRESET

Settings

Show Diff

Only show mismatching lines

Show Prediction

SAVE RESULT

LOAD RESULT

OCR Tools

eScriptorium:

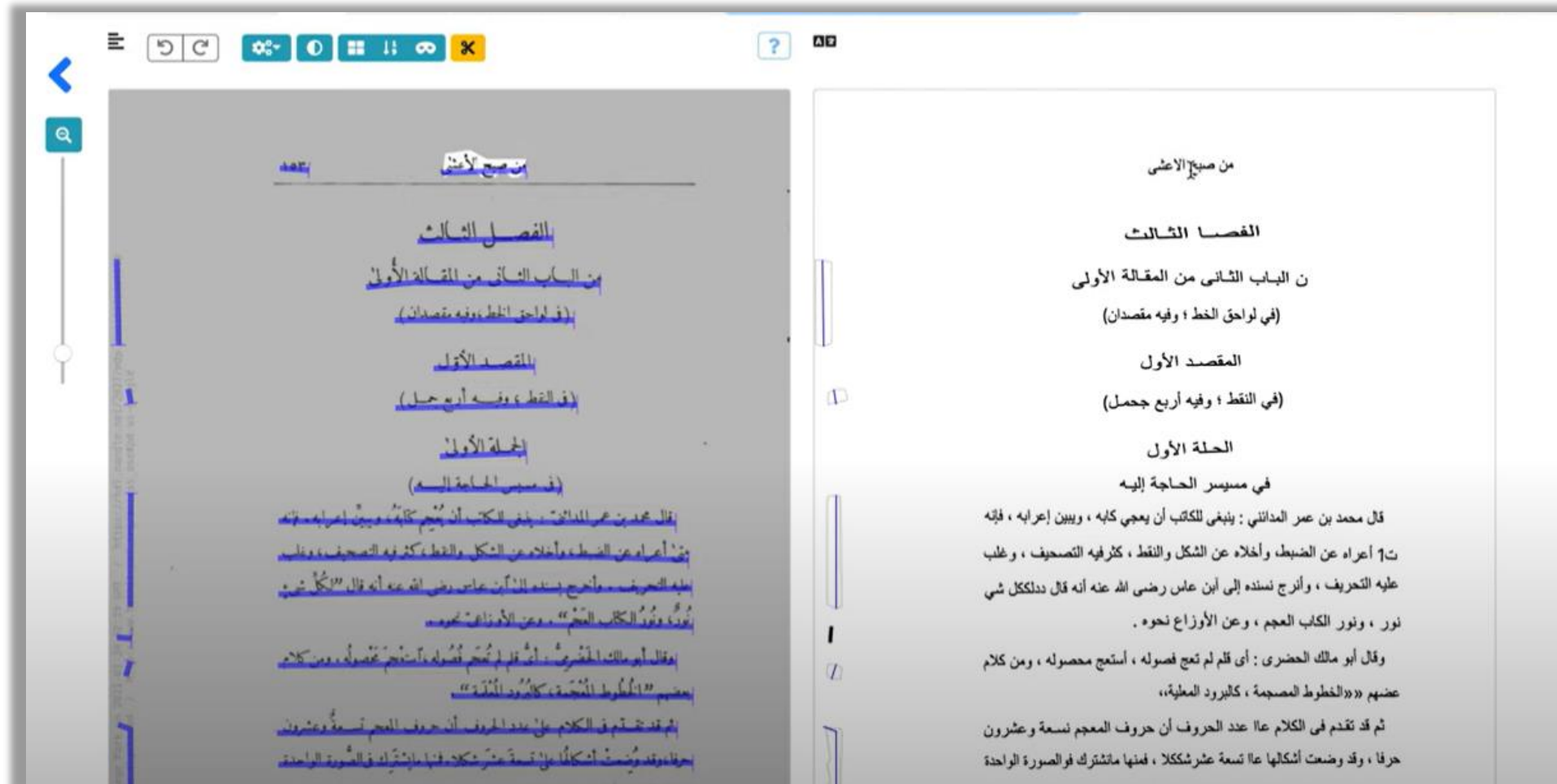


Image: Youtube eScriptorium Tutorial Part I: the Basics by OpenITI

Periodicals



Suḥan

Yağmā

Farhang-i Īrān-zamīn

Rāhnama-i kitāb

Bar-rasīhā-i tāriḥī

Full-Text Recognition

Transkribus

- Previously used Tesseract for open source OCR engine for Arabic and Persian
- Character accuracy (Cs) of 69.28% and a letter accuracy (Ls) of 70.19% over the evaluation corpus with approx. 180 groundtruth pages.
- A second OCRing of the Rahbar material with this improved data will take place in Q3 2024

GroundTruth Data for Arabic and Persian Texts

- The text is faithfully reproduced
- Printing errors are not corrected
- Spelling errors in the original will not be corrected
- Vocalization characters and other diacritics are formed by combining the Unicode characters
- Punctuation marks are rendered as indicated in the text
- Words are separated by spaces
- Ligatures with the help of Unicodes
- Save in ALTO XML format with UTF-8 encoding

OCR Metrics

Ca (character accuracy - all visible characters and spaces)

Ls (letter accuracy - characters without spaces, punctuation and numbers)

Wa (word accuracy - referring to "words", i.e. tokens separated by a space)

BoW (bag of words - "words" without regard to the order)

مردمان با ذوق و استعداد بادییات میبرد و داختمند از این جهت از مشربان ادب کاسته کتر شاعر و سخن سنج برینرد بیاید باید دید که در ایران هم این طوور است؟ محققاً نه کجا هستند فضلاي ما و چه آثاری از آنان بجااست کدام علم و هنر توجه مردم را بخود کشیده است باز مثل سابق چشم همه دوخته بادییات است شعرگو در ایران کم نیست بلکه در میان آنان فضل و علم کم است که آنانرا ترقی داده بزرگت کند بنا بر این اسباب انحطاط ادب در ایران ترقی علم و معرفت و صنایع و هنر نیست بلکه ترقی جهل و نادانی است برای آنکه از طبایع سرشار استفاده شود و منظومات از سستی و خالی بیرون آید و فکرها بلند و اندیشه ها سنجیده گردد و از بر تو آن روح نازه و خری مردم دمیده شود گذشته از علم عروض و اطلاق از اشعار متقدمین علوم دیگر هم لازم است متقدمین غالباً بعلوم عصر خود از قبیل علم لغت و تاریخ و حکمت و نجوم آشنا بودند ما در این قرون هیچ کدام را لازم نمیدانیم بنظر ما فقط از متقدمین سرقت معانی و فکر و تمییرات لازم است از این جهت است که ذوق سلیم ایرانیان که عادت بخواندن اشعار لطیف و شنیدن مضامین نغز و شیرین متقدمین دارند منظومات ما را نیز لازم ندانسته زیادتی شمرده از کوشی شنیده از گوش دیگر رد میکنند و بلکه کله شاعر را در این قرون اخیر لفظ مترادف چرسی و تریاکی و الکلی و هرزه و مسخره میدانند

بسی از ما رومه شدن ایران بلك روح مگس و باک نشسته تجددی بصرای ایران دمیده شد بطوریکه موقتاً دست غارت از خوان معزّی و سعدی کشیدند ولی علم که ممد نشو و نمای ایران نوزاد بود نقصان نمود لاجرم نونهال بشر نرسید بهای خای و طفولیت خویش ماند تا بتدریج خشک شد ولی اعید میتوان داشت که دگر باره دماغهای افسرده ترو نازه کشته خیالات بدیع ایجاد کند هر روز که علم و معرفت در ایران زیاد شد و ملیت ایران قوت گرفت نظم و نثر نیز روان نوي گرفته اثرات خوبی خواهد بخشید جنساً ایرانیان اهل ذوق میباشند و زبان فارسی دریای بیکرانی است از لغات و تعبییرات و اصطلاحات و از شعرای متقدم که در هر رشته از رزم و

مردمان با ذوق و استعداد بادیات میپرداختند از این جهت از مشتریان ادب کا سته کمتر شاعر و سخن سنج بوجود میآید باید دید که در ایران هم این طور است؟ محققاً نه کجا هستند فضایی ما و چه آثاری از آنان بجاست کدام علم و هنر توجه مردم را بخود کشیده است باز مثل سابق چشم همه دوخته بادیات است شعرگو در ایران کم نیست بلکه در میان آنان فضل و علم کم است که آنانرا ترقی داده بزرگ کند بنابر این اسباب انحطاط ادب در ایران ترقی علم و معرفت و صنایع و هنر نیست بلکه ترقی جهل و نادانی است برای آنکه از طبایع سرشار استفاده شود و منظومات از سُستی و خامی بیرون آید و فکرها بلند و اندیشه‌ها سنجیده گردد و از برتو آن روح تازه و خرمی بمردم دمیده شود گذشته از علم عروض و اطلاع از اشعار متقدمین علوم دیگر هم لازم است متقدمین غالباً بعلوم عصر خود از قبیل علم لغت و تاریخ و حکمت و نجوم آشنا بودند ما در این قرون هیچ کدام را لازم نمیدانیم بنظر ما فقط از متقدمین سرقت معانی و فکر و تعبیرات لازم است از این جهت است که ذوق سلیم ایرانیان که عادت بخواندن اشعار لطیف و شنیدن مضامین نغز و شیرین متقدمین دارند منظومات ما را لازم ندانسته زیادتی شمرده از گوشه شنیده از گوش دیگر رد میکنند و بلکه کلمه شاعر را در این قرون اخیر لفظ مترادف چرسی و تریاک والکلی و هرزه و مسخره میدانند

پس از مشروطه شدن ایران بک روح ملی و بک نفخه تجدیدی بشعراى ایران دمیده شد بطوریکه موقتاً دست غارت از خوان مُعزّی و سعدي کشیدند ولی علم که ممد نشو ونمای این نوزاد بود نقصان نمود لاجرم نونهال بثمر نرسید بهمان خامی و طفولیت خویش ماند تا بتدریج خشک شد ولی امید میتوان داشت که دگر باره دماغهای افسرده ترو تازه کشته خیالات بدیع ایجاد کند هر روز که علم و معرفت در ایران زیاد شد و ملیت ایران قوت گرفت نظم و نثر نیز روان نوي گرفته اثرات خوبی خواهد بخشید جنساً ایرانیان اهل ذوق میباشند و زبان فارسي دریاي بیکرانی است از لغات و تعبیرات و اصطلاحات و از شعراى متقدم که در هر رشته از رزم و

بیاچه

مردمان با ذوق استعداد بادیات میپرداختند از این جهت از مشتریان ادب کا سته سرشانی سخن سنج بوجود میآید باید دید که در ایران این طور است؟ محققاً نه کجا هستند فضایی ما و چه آثاری از آنان بجاست کدام علم و هنر توجه مردم را بخود کشیده است باز مثل سابق چشم همه دوخته بادیات است شعرگو در ایران کم نیست بلکه در میان آنان فضل و علم کم است که آنانرا ترقی داده

درت کند بنابر این اسباب انحطاط ادب در ایران ترقی علم معرفت صنایع و هنر نیست بلکه ترقی جهل نادانی است برای آنکه از طبایع سرشار استفاده شود منظومات از سستی خامی بیرون آید فکرها بلند و اندیشه‌ها

سنجیده ترند از برتو آن روح تازه خرمی بمرم دمیده شود گذشته از علم عروض اطلاع از اشعار متقدمین علوم دیگر هم لازم است متقدمین غالباً بعلوم عصر خود از قبیل علم لغت تاریخ حکمت نجوم آشنا بودند ما در این قرون هیچ کدام را لازم نمیدانیم بنظر ما فقط از متقدمین سرقت معانی و فکر و

تعبیرات لازم است از این جهت است که ذوق سلیم ایرانیان که عادت بخواندن اشعار لطیف شنیدن مضامین نغز شیرین متقدمین دارند منظومات ما را نیز لازم ندانسته زیادتی شمرده از گوشه شنیده از گوش دیگر رد میکنند بلکه کلمه شاعر را در این قرون اخیر لفظ مترادف چرسی تریاک والکلی

هرزه مسخره میدانند

پس از مشروطه شدن ایران بک روح ملی و بک نفخه تجدیدی بشعراى ایران دمیده شد بطوریکه موقتاً دست غارت از خوان مُعزّی و سعدي کشیدند

ولی علم که ممد نشو ونمای این نوزاد بود نقصان نمود لاجرم نونهال بثمر نرسید بهمان خامی و طفولیت خویش ماند تا بتدریج خشک شد ولی امید میتوان داشت که دگر باره دماغهای افسرده ترو تازه

کشته خیالات بدیع ایجاد کند هر روز که علم و معرفت در ایران زیاد شد و ملیت ایران قوت گرفت نظم و نثر نیز روان نوي گرفته اثرات خوبی خواهد بخشید جنساً ایرانیان اهل ذوق میباشند و زبان فارسي دریاي بیکرانی است از لغات و تعبیرات و اصطلاحات و از شعراى متقدم که در هر رشته از رزم و

دیباچه

مردمان با ذوق و استعداد بادیات میپرداختند از این جهت از مشتریان ادب کا سته کمتر شاعر و سخن سنج بوجود میآید باید دید که در ایران هم این طور است؟ محققاً نه کجا هستند فضایی ما و چه آثاری از آنان بجاست کدام علم و هنر توجه مردم را بخود کشیده است باز مثل سابق چشم همه دوخته بادیات است شعرگو در ایران کم نیست بلکه در میان آنان فضل و علم کم است که آنانرا ترقی داده بزرگ کند بنابر این اسباب انحطاط ادب در ایران ترقی علم و معرفت و صنایع و هنر نیست بلکه ترقی جهل و نادانی است برای آنکه از طبایع سرشار استفاده شود و منظومات از سُستی و خامی بیرون آید و فکرها بلند و اندیشه‌ها سنجیده گردد و از برتو آن روح تازه و خرمی بمرم دمیده شود گذشته از علم عروض و اطلاع از اشعار متقدمین علوم دیگر هم لازم است متقدمین غالباً بعلوم عصر خود از قبیل علم لغت و تاریخ و حکمت و نجوم آشنا بودند ما در این قرون هیچ کدام را لازم نمیدانیم بنظر ما فقط از متقدمین سرقت معانی و فکر و تعبیرات لازم است از این جهت است که ذوق سلیم ایرانیان که عادت بخواندن اشعار لطیف و شنیدن مضامین نغز و شیرین متقدمین دارند منظومات ما را نیز لازم ندانسته زیادتی شمرده از گوشه شنیده از گوش دیگر رد میکنند و بلکه کلمه شاعر را در این قرون اخیر لفظ مترادف چرسی و تریاک والکلی و هرزه و مسخره میدانند

پس از مشروطه شدن ایران بک روح ملی و بک نفخه تجدیدی بشعراى ایران دمیده شد بطوریکه موقتاً دست غارت از خوان مُعزّی و سعدي کشیدند ولی علم که ممد نشو ونمای این نوزاد بود نقصان نمود لاجرم نونهال بثمر نرسید بهمان خامی و طفولیت خویش ماند تا بتدریج خشک شد ولی امید میتوان داشت که دگر باره دماغهای افسرده ترو تازه کشته خیالات بدیع ایجاد کند هر روز که علم و معرفت در ایران زیاد شد و ملیت ایران قوت گرفت نظم و نثر نیز روان نوي گرفته اثرات خوبی خواهد بخشید جنساً ایرانیان اهل ذوق میباشند و زبان فارسي دریاي بیکرانی است از لغات و تعبیرات و اصطلاحات و از شعراى متقدم که در هر رشته از رزم و

Text Recognition

از جناب رضا علیه السلام سؤال نمود که دلیل بر خلافت جد شما چیست حضرت
فرمود آیه انفسنا مومن گفت لولا نسائنا حضرت فرمود لولا ابنائنا این
سؤال و جواب را تطبیق بکن و مقصود را بیان نما ساعتی تأمل نموده

از جناب رضا علیه السلام سؤال نمود که دلیل بر خلافت جد شما چیست حضرت
فرمود آیه انفسنا مومن گفت لولا نسائنا حضرت فرمود لولا ابنائنا این
سؤال جواب را تطبیق بکن مقصود را بیان نما ساعتی تأمل نموده

Data analysis

[1767907966_00000049]

[Cs:84.90(1497), Ls:87.46(1180)(+ 2.56), Ws:30.69 (290), BoWs:39.24 (290)(+ 8.55)]

Text Recognition

Marking the text and line region

یگرد تازگی ند ارد • شما بعنوان نماینده قانون بنام ملت، آزادی،
راه پشت میز محاکمه جلب کرده اید که بنابه تشخیص آقای
، بیگانه پرستی، داشتن افکار مضرو خطرناک حتی باتهام سو
ی، زهی عدالت • آقای دادستان بنام قانون و بنام عدالت برای
حکم اعدام نموده اند • باور کنید همانقدر که شما چه در موقع درخواست
آرام و بی تفاوت هستید من نیز بهمان اندازه آرام و بی اعتنا هستم
مقصد بد انم • زندگی برای من مدتهاست که دیگر دور من و آنچه
، وجود من ، خواسته های من، آرمانها و امید های من، دلهره ها
• ملتم، خواسته های ملتم، آرمانها و امیدها، دلهره ها و دلخوشی-
• اگر زندگی بایستی عبارت از خوردن چند لش اضافی گوسفند و
نبودنش هزار بار از بودنش بهتر است • محرک من در بیان آخرین

4-4 دادستان و سازمان امنیت بجا سوسی، بیگانه پرستی، داشتن افکار مضرو خطرناک حتی باتهام سو
4-5 قصد متهمند. زهی حقیقت، زهی آزادی، زهی عدالت. آقای دادستان بنام قانون و بنام عدالت برای
4-6 من و دوستم از دادگاه تقاضای صدور حکم اعدام نموده اند. باور کنید همانقدر که شما چه در موقع درخواست
4-7 اینگونه احکام و چه در موقع صدور آنها آرام و بی تفاوت هستید من نیز بهمان اندازه آرام و بی اعتنا هستم.
4-8 آنگاه دل من میتواند بلرزد که خود را مقصد بدانم زندگی برای من مدتهاست که دیگر دور من و آنچه
4-9 تنها مربوط بشخص من است دور نمیزند، وجود من، خواسته های من، آرمانها و امیدهای من، دلهره ها
4-10 و دلخوشی های من همه و همه در وجود ملتم، خواسته های ملتم، آرمانها و امیدها، دلهره ها و دلخوشی-
4-11

Full-Text Presentation

-۱۶-

برخی هوش تند و زیرکی برخی جانوران را به رخ ما می کشیدند .
پاسخ دادیم : هوش چیز دیگر است و آنچه ما می گوئیم چیز دیگر . بسیاری
از جانوران دریافتهایی می دارند که در آدمی نیست و در همانحال از دلسوزی
بیگدیگر و مانند اینها بی بهره می باشند. مثلاً سگ هوش شگفت انگیزی میدارد
و امروز چه در جنگ و چه در کارهای پلیسی از هوش او سودجویی میشود .
با اینحال در سگ کشیها می بینیم یکی را که زهر خورانیده اند و در غلطیده
جان می کند سگهای دیگر در پهلوی آن بی هیچ اندوه دلسوزی می ایستند.

-۱۶-

برخی هوش تندی برخی جانوران را به رخ مکتشیدند
پاسخ دادیم هوش چیز دیگر است آنچه ما می گوئیم چیز دیگر بسیاری
از جانوران دریافتهایی میدارند که در آدمی نیست در همانحال از دلسوزی
بیگدیگر و مانند اینها بی بهره میباشند مثلاً سگ هوش شگفت انگیزی میدارد
امروز چه در جنگ چه در کارهای پلیسی از هوش او سودجویی میشود
با اینحال در سگ کشیها بینیم یکی را که زهر خورانیده اند در غلطیده
جان میکند سگهای دیگر در پهلوی آن بی هیچ اندوه دلسوزی می ایستند

Data analysis: [1767910290_00000018] [Cs:17.93(1840), Ls:14.54(1417)(- 3.39)]

-۱۰-

شرمی بخود راه نمیداد . دیگری که او نیز استاد است بهر کار زشت و پستی
میپرداخت . چون با کسانی کشاکش های حزبی میداشت روزی دیدم دروغهایی
بآنان می بست و چنبن می گفت : «در مبارزه آدم باید از هیچ کاری باز نایستد .
اساس زندگانی مبارزه است» .

کسانی را می شناختم که وزیر می بودند و بانداشتن نیاز رشوه می گرفتند
و دزدی می کردند و عنوانشان این می بود : «زندگانی مبارزه است» . شنیدم
یکی گفته بوده : «من اگر ندزدم بد کرده ام . برخلاف طبیعت رفتار نموده ام» .
شنیدنیست که بعضی از ملایان این بدآموزیها را یاد گرفته و در آن

-۱۰-

شرمی بخود راه نمیداد دیگری که او نیز استاد است بهر کار زشتی
میپرداخت چون با کسانی کشاکش های حزبی میداشت روزی دیدم دروغهایی
بآنان می بست چنبن میگفت در «مبارزه آدم باید از هیچکاری باز نایستد
اساس زندگانی مبارزه است»

کسانی را می شناختم که وزیر می بودند بانداشتن نیاز رشوه می گرفتند
نزد می کردند و عنوانشان این می بود : «زندگانی مبارزه است» شنیدم
یکی گفته بوده : «من اگر ندزدم بکرده ام برخلاف طبیعت رفتار .» نموده ام
شنیدنیست که بعضی از ملایان این بدآموزیها را یاد گرفته در آن

Data analysis: [1767910290_00000012] [Cs:96.89(1798), Ls:97.66(1370)(+ 0.77)]

Github Issue

ULB
SACHSEN-ANHALT

Universitäts- und Landesbibliothek Sachsen-Anhalt
3 followers · Germany, Halle (Saale) · <https://bibliothek.uni-halle.de/>

Overview · Repositories 34 · Projects 2 · Packages · People 6

Popular repositories

- ulb-zeitungsprojekt-hp1**
Training data from "Hauptphase I" of project "Digitalisierung deutscher Zeitungen"
Makefile · 11 stars · 1 fork
- ocr-pipeline**
OCR Pipeline Module for Digitalization Workflows
Python · 5 stars · 1 fork
- mets-model**
Forked from [MyCoRe-Org/mets-model](#)
Simple METS Api
Java · 2 stars

right to left and left to right Orientation of the Persian digital copies in the full-text display on IIF, DFG Viewer and PDF #54 New issue

Open · MaidaButtar opened this issue on Nov 9, 2023 · 8 comments

MaidaButtar commented on Nov 9, 2023

The following problems occur when recognizing and displaying the left-to-right and right-to-left orientation in the full-text display of the IIF and DFG Viewer and in the PDF files in Persian:

IIF Viewer: The order of the words is correct, but the letters in the words are reversed. The order of digits is correct (reason: numbers are read from left to right)

DFG Viewer: Line breaks are all gone, the order of words is halfway correct, but again, the letters in the word are reversed. The order of digits is correct.

PDF: Order of words reversed at line level, but letters in the word are not reversed. Order of digits correct..

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone

Fachinformationsdienst Nahost-, Nordafrika- und Islamstudien

خدمة تقديم المعلومات المختصة بدراسة الشرق الاوسط وشمال افريقيا والدراسات الاسلامية

Orta Doğu, Kuzey Afrika ve İslam Araştırmaları Bilgi Hizmetleri

اطلاعات و خدمات کارشناسی در زمینه مطالعات خاورمیانه و شمال افریقا و پژوهشهای اسلامی

Specialised Information Service Middle East, North Africa and Islamic Studies



Thank you for your attention!

Contact: maida-farooq.buttar@bibliothek.uni-halle.de
www.menalib.de



UNIVERSITÄTS- UND LANDESBIBLIOTHEK
SACHSEN-ANHALT

DFG Deutsche
Forschungsgemeinschaft