



what is Decapod?

Decapod is a project focused on building a low-cost digitization solution that will allow for rare materials, materials held in collections without large budgets, and other scholarly content to be digitized into a high-quality PDF format. This project will work to incorporate the hardware and software necessary to accomplish this goal.

who is Decapod?

Technical University of Kaiserslautern (DFKI)
Image Understanding and Pattern Recognition
(IUPR)

University of Toronto (UTor)
Adaptive Technology Resource Centre (ATRC)
Fluid Project

JSTOR

Funded by a grant from
The
Andrew W. Mellon
Foundation

Fluid

problems we're addressing

cost (equipment and staff and training)

errors (process is imperfect)

fragile materials

open source (make it available and flexible)

challenges with OCR

fonts and typefaces

handwriting

internationalization (different languages and alphabets)

complex layouts (multi column; illustrations and

complex lists and tables)

old, fragile, discoloured texts

mobile

web

tagging

search

OCROPUS

Technical University of Kaiserslautern (DFKI)
Image Understanding and Pattern Recognition
(IUPR)

Google
high quality / low cost

built on top of HPs Tesseract OCR Engine

OCROpus Processing Pipeline



Book Scanning Pipeline



OCROPUS



(a)



(b)



(c)

what it will do

1. Allow the camera based capture of bound material by using computer vision techniques to produce flat, clean page images equivalent to those produced from a flat bed scanner.
2. Remove the need for extensive operator intervention in the capture process by detecting scan problems and allowing the operator to rectify the scan immediately.
3. Reduce user intervention in the conversion process by using advanced document understanding techniques to remove almost all intervention, and by reducing the remainder to very simple "1-click" operations.

what it will do

4. PDF/A outputs will be visually faithful to the original, searchable, and widely usable.

5. Allow the output to be viewable on mobile devices that support PDF reflow.

6. Remove the need for deep software, hardware or digitization skills by integrating all software components into a turnkey end-to-end solution.

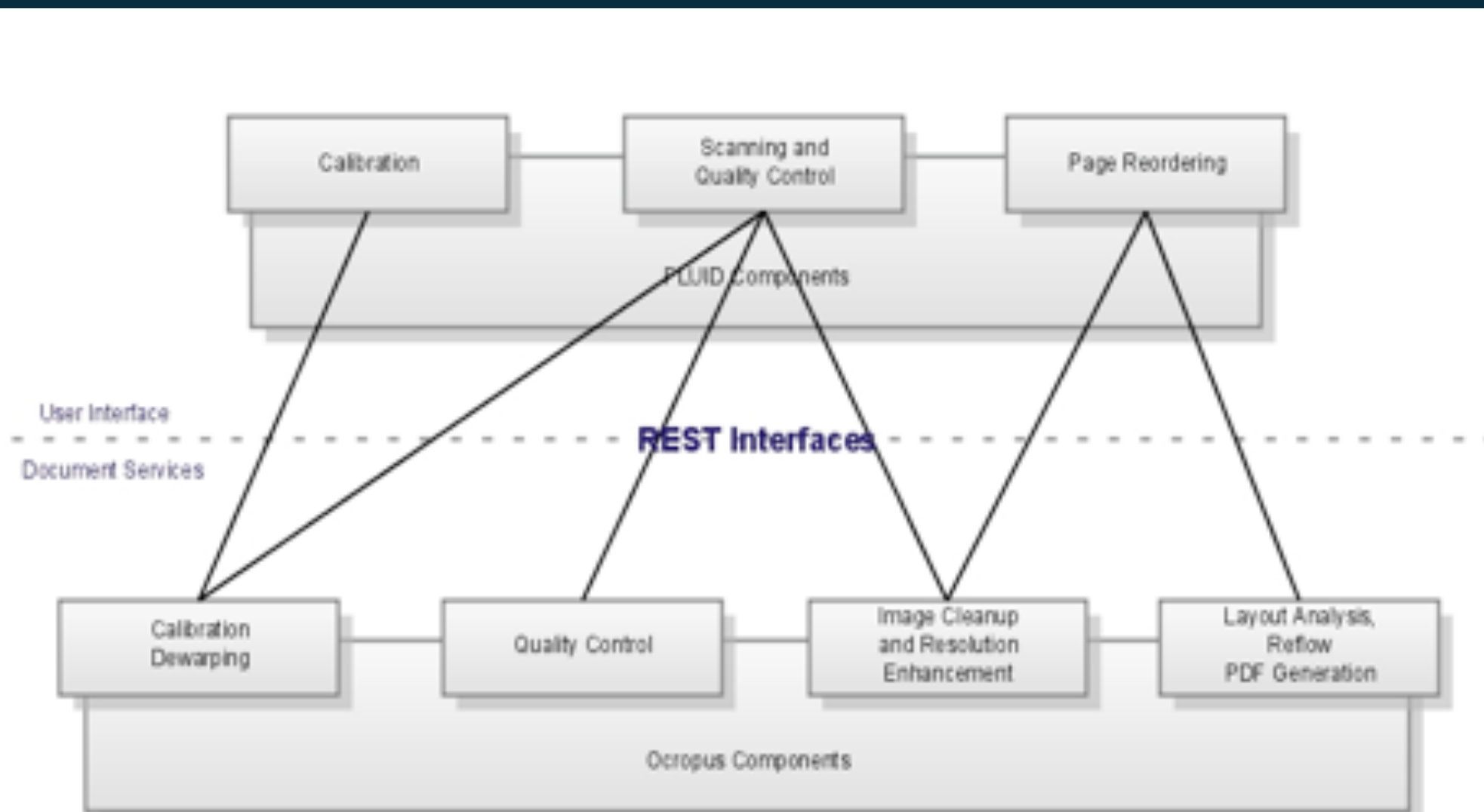
7. Remove capital cost barriers by using consumer grade cameras.

8. Reduce operational cost barriers by allowing volunteers or local staff to operate the system with minimal training or commitment.

OCROPUS architecture

pluggable:

layout analysis, character recognition, and language modeling -- meant for high volume




The Solution




The Interface

The screenshot displays the 'Book Capture' web application interface. At the top left, the logo 'Book Capture' is shown with a stylized bird icon. In the top right corner, the user's name 'Mary Taylor' and a 'Sign Out' link are visible. Below the header, a navigation bar shows the book title 'The Secret Garden' and a close button. The main area is a grid of scanned pages. The first page is the book cover, and the second page is a table of contents. The third page is highlighted with a blue border and labeled 'Page 2'. The left sidebar contains two sections: 'Tools' and 'View'. The 'Tools' section includes radio buttons for 'Flatten page', 'Correct lighting', 'Remove noise', 'Crop white space', 'Brightness', and 'Contrast', along with sliders for 'Brightness' and 'Contrast', and an 'Apply' button. The 'View' section includes 'Zoom' controls (two magnifying glass icons), 'View' controls (two document icons), and 'Orientation' controls (two document icons). The bottom right corner features the 'Fluid' logo, which consists of a purple asterisk-like shape above the word 'Fluid' in a green font.

Tools 



The Secret Garden - Page 1,2 



- Flatten page
- Correct lighting
- Remove noise
- Crop white space
- Brightness
- Contrast







[Apply](#)

View 

Zoom:  

View:  

Orientation:  

Tag 

page 2, chapter 1

1

CHAPTER I

THERE IS NO ONE LEFT

When Mary Lennox was sent to Misselthwaite Manor to live with her uncle everybody said she was the most disagreeable-looking child ever seen. It was true, too. She had a little thin face and a little thin body, thin light hair and a sour expression. Her hair was yellow, and her face was yellow because she had been born in India and had always been ill in one way or another. Her father had held a position under the English Government and had always been busy and ill himself, and her mother had been a great beauty who cared only to go to parties and amuse herself with gay people. She had not wanted a little girl at all, and when Mary was born she handed her over to the care of an Ayah, who was made to understand that if she wished to please the Mem Sahib she must keep the child out of sight as much as possible. So when she was a sickly, fretful, ugly little baby she was kept out of the way, and when she became a sickly, fretful, toddling thing she was kept out of the way also. She never remembered seeing familiarly anything but the dark faces of her Ayah and the other native servants, and as they always obeyed her and gave her her own way in everything, because the Mem Sahib would be angry if she was disturbed by her crying, by

Copyrighted material

2

the time she was six years old she was as tyrannical and selfish a little pig as ever lived. The young English governess who came to teach her to read and write disliked her so much that she gave up her place in three months, and when other governesses came to try to fill it they always went away in a shorter time than the first one. So if Mary had not chosen to really want to know how to read books she would never have learned her letters at all.

One frightfully hot morning, when she was about nine years old, she awakened feeling very cross, and she became crosser still when she saw that the servant who stood by her bedside was not her Ayah.

"Why did you come?" she said to the strange woman. "I will not let you stay. Send my Ayah to me."

The woman looked frightened, but she only stammered that the Ayah could not come and when Mary threw herself into a passion and beat and kicked her, she looked only more frightened and repeated that it was not possible for the Ayah to come to Missie Sahib.

There was something mysterious in the air that morning. Nothing was done in its regular order and several of the native servants seemed missing, while those whom Mary saw slunk or hurried about with ashy and scared faces. But

Copyrighted material

Go to page:

Delete selected page

Tools

- Flatten page
- Correct lighting
- Remove noise
- Crop white space
- Brightness



- Contrast



Apply

Scanning: The Secret Garden - Page 2

View

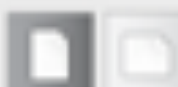
Zoom:



View:



Orientation:



114

International Journal of Ethics.

his "Republic," with its elaborate scheme of social relations, before he can give an answer that is more than a mere empty form. When the answer, Justice is a giving every one his own, is offered, it appears entirely satisfactory, but, on closer scrutiny, turns out to be entirely untrue, until some method is discovered of determining what is every one's own. So, in the present case, to answer the question, What is the aim of education? by saying, To make good men and women, or good citizens, does not help us, unless there can be found some method of determining what is meant by "good" in such connection, and this involves a theory or conception of the universe and of man's place in it.

Now, while we may lament that the absence of any such conception leaves our present education in the chaotic condition in which we find it, we can really see how such a state of things has come about. The old medieval conception, based upon supernatural revelation, or what was considered such, is passing away, and being its hold upon practice, and so new conceptions, based upon science, as all future conceptions must be, has yet arisen to take its place. We have, no doubt, certain convincing upon schools-theories of an a priori sort; but, as they do not correspond to the facts of the world, or even to universal intelligence, they are of no permanent value. What we most need to-day is a co-operative effort on the part of a little group of earnest men, well versed in the different sciences, physical and spiritual, to present such a scheme of the universe as shall form the basis of a rational, saving education, leaving room, of course, for continual additions, modifications, and improvements. While we are anxiously waiting for this, it cannot be but interesting and instructive to consider the system of the "Brethren of Purity," put forward with this view, and based on such science as existed in the tenth century in the most cultivated circles.

This system is laid down in a Cylogordia, which must have been written about the year 1000 A.D., and which was printed in Calcutta, for the first time, in 1812, and again in 1842, in four large volumes of densely crowded, unvocalized Arabic.

relevance

can you use it?

how can you contribute to it?

Links

- <http://sites.google.com/site/decapodproject/Home>
- <http://en.wikipedia.org/wiki/OCRopus>
- <http://sites.google.com/site/ocropus/>
- <http://fluidproject.org/projects/decapod/>

Links

- http://www.economist.com/science/tq/displaystory.cfm?STORY_ID=13174399
- openlibrary.org (an Internet Archive project)