DIVA Services - Spotlight

Experimenting with Document Image Analysis Methods on the Web

Marcel Würsch (@lunactic), Michael Bärtschi, Rolf Ingold, and Marcus Liwicki
DIVA Group, University of Fribourg, Switzerland
Our Vision

Make DIA methods easy available

Accessible over the internet

Hosted on a «Cloud» infrastructure

DIVA Services-Spotlight is one frontend
DIVAServices – The Power Behind

RESTFul Web Service

Using HTTP commands

No computation on the client

No local installation needed

No domain knowledge needed
Workflow When Using DIVAServices
18 Methods are Available on DIVAServices

Original

Ocropus

Image Enhancement

Text Line Extraction

Original

Artificial Degradation

Layout Analysis

Color Inverting
DIVA Services - Spotlight
<table>
<thead>
<tr>
<th>It is</th>
<th>It is not</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Simple Web-Interface</td>
<td>Production ready</td>
</tr>
<tr>
<td>Testing DIA possibilities</td>
<td>For large scale experiments</td>
</tr>
<tr>
<td>Finding good parameters</td>
<td>A specialized tool</td>
</tr>
</tbody>
</table>
DIVA Services-Spotlight – Features

Upload your images
Deleted after closing the browser

Test all available methods

One possible visualization of results
Live Demonstration
What if I Have a Specific Use Case?

Interact with DIVA Services programmatically
Java library available
Python library by 2017

Build user interfaces for your needs
From DIVA Services to TEI

Our own use case

Goal: Create transcriptions

Workflow

Manage Images
Create Zones on Images
Extract Text Lines
Save to TEI
//get the selected region
Rectangle r = (Rectangle) selectedZone.getArea().getShape();

//Create the communication class
DivaServicesCommunicator communicator = new DivaServicesCommunicator("http://divaservices.unifr.ch");

//Call the extraction method
DivaServicesResponse<Polygon> resp = communicator.runSeamCarvingTextlineExtraction(bi, r, 0.0003f, 3.0f, 4, false);

//Retrieve the results
List<Polygon> polygons = resp.getHighlighter().getData();
DIVAServices and DIVAServices-Spotlight

LGPL v2.1

Image Data
Stored on our infrastructure
Needs to be Creative Common

Methods
Sometimes not Open Source
Conclusion and Outlook

**DIVA Services**
Makes methods available
Interact with any programming language

**DIVA Services-Spotlight**
Experiment with the available method
Find parameters
Not for production

**Future**
Tool for providing methods (soon)
Workshop at DH2017
Thank You For Your Attention

Contact: marcel.wuersch@unifr.ch / @lunactic

DivaServices-Spotlight
http://divaservices.unifr.ch/spotlight

Project Website