Comparison of optical character recognition (OCR) software

by

Angelica Gabasio

Department of Computer Science
Lund University

June 2013

Master's thesis work carried out at *Sweco Position*.

Supervisors	Björn Harrtell, Sweco
	Tobias Lennartsson, Sweco
Examiner	Jacek Malec, Lund University

Abstract

Optical character recognition (OCR) can save a lot of time and work when information stored in paper form is going to be digitized, at least if the output from the software is accurate. If the output is very inaccurate, there will be need for a lot of post-processing, and correcting all of the errors might result in more work than manually typing in the information.

The purpose of the thesis is to run different OCR software and compare their output to the correct text to find out which software generates the most accurate result. A percentage value of the error is calculated to find out how accurate each software is. Different kinds of images will be tested to find out how accurate the software are on images of different quality.

Both commercial and open source software has been used in the thesis to find out if there is a difference.

Keywords: OCR, comparison, Tesseract, Ocrad, CuneiForm, GOCR, OCRopus, TOCR, Abbyy CLI OCR, Leadtools OCR SDK, OCR API Service, Wagner-Fischer algorithm

Acknowledgements

Thanks to Björn Harrtell and Tobias Lennartsson, my supervisors at Sweco Position, for providing feedback and guidance during the thesis work.

Also thanks to everyone else at Sweco Position for being so supportive and interested in the thesis progress.

Thanks to my examiner, Jacek Malec, for feedback on this report.

Contents

1 Introduction

Optical character recognition (OCR) is a way of extracting plain text from images containing text. These images can be books, scanned documents, handwritten text etc. The purpose of OCR can be to be able to search the text that has been scanned, to automatically process filled forms, or to digitize information stored in books.[1] Google books (`http://books.google.com/`) is an example of the usage of OCR scanning, where books are scanned and OCR is applied to make the text searchable.[2]

A lot of text and information is stored in paper form only, but since our surroundings is getting more and more digitized, the information would be more accessible if it was stored in digital form as well. OCR scanning can be a way to decrease the amount of manual work needed to digitize the information.

For OCR scanning to be useful it is important that the output is correct, or almost correct. The output from different software may differ a lot even if they are applied on the same image. If the output is very inaccurate there may be need for a lot of manual post-processing, and in some cases the outputted text might not be readable at all. In this thesis a comparison of different OCR tools will be made by comparing their output text to the correct text. The comparison will be done using a string comparison algorithm[4] that returns a number indicating how different two strings are. This number is used to calculate a percentage value of the output error.

1.1 History of OCR

The development of OCR scanning started in the 1950s, but there had been some similar work done earlier. In the 1960s, OCR got more widespread and the development of OCR machines increased, which lead to better accuracy of the OCR scan.

In the beginning, OCR scanning was very slow, there was a machine that read upper case alphanumeric characters at the speed of one character per minute. In the late 1950s there existed machines that could read both upper and lower case letters.[3]

In the late 1960s, there were OCR machines that could use machine learning to be able to recognize more fonts. Solutions that made it possible for the OCR machines to read characters that were imperfect, was developed at this time as well.[3]

One of the first applications of early OCR technology was the *Optophone* that was developed in 1912. It was a handheld machine that was moved over a page, and outputted different tones corresponding to the different

characters. It was developed as a help for people with vision problems, but it did not get very widespread since it required a lot from the user to learn to recognize the different tones.[3]

The first use of a commercial OCR machine in a business helped reduce the processing times of documents from one month to approximately one day, this was in 1954. At this time, OCR was also used as a help for sorting mail in post offices.[3]

One of the first OCR developers suggested the use of a standard font to help the OCR recognition in the 1950s, and in the 1960s two new fonts were introduced for this purpose. The first one, OCR-A, was designed in America, and included upper case letters, numerals and 33 other symbols. Three years later, lower case letters were added to the font as well. Europe wanted another look on the standardized font, so they designed their own standard font, OCR-B.[3]

In the beginning, the paper that was going to be OCR scanned had to be moved to scan different parts of it. The machines could do this by themselves, or either the paper or the scanner could be moved manually.[3]

2 Theory

The basic algorithm of OCR scanning consists of the following steps[1, 2]:

- Preprocessing
- Layout analysis
- Character recognition
- Output

In the *preprocessing* step the image is modified in a way that will make the OCR algorithm as successful as possible. The preprocessing usually includes:

1. Converting the image to only black and white pixels. This is called binarisation.
2. Removing noise from the binarised image.
3. Rotating the image to align the text as horizontal as possible.

In addition to these modifications there are several other less common steps that can be used for preprocessing if needed.[2] The preprocessing can be done either internally in the software or manually before the OCR scan.

The purpose of the *layout analysis* is to determine how the text should be read. This can include identifying columns or pictures in the text.[2]

After the analysis it is time for the *character recognition*, which is usually applied on one line of text at a time. The algorithm breaks the line into words, and then the words into characters that can be recognized.[2]

The algorithm uses a database of letters, numerals and symbols to match the characters. The algorithm will calculate a matching value for each character and selects the option with the best value.[1]

A way of improving the output is to check if the scanned word exists in a dictionary. If it doesn't, it is likely that some of the analyzed letters are wrong, and some of the letters may be changed so the word matches a word in the dictionary. This is done by adding a penalty value to the matching value if the analyzed character result in a word not in the dictionary. This makes it possible for another letter to get a better matching value if the word with the new letter is part of the dictionary. The letters are not always changed, sometimes a non-dictionary word gets a better matching score, even with the penalty value, than the one with the changed letters that is in the dictionary. Even if using a dictionary may help the output accuracy it is no guarantee that the correct word will be found.[1] An example of the use of a dictionary can be seen in Figure 1.

Other ways to get a better output is to use machine learning, where the OCR tool can be trained to recognize more languages or fonts, or to use the knowledge of which character combinations are more likely to appear.

The better the image is, the more accurate the result gets. If the OCR algorithm is applied on a horizontal text with very little or no noise and the letters are well separated from each other it is possible to get a very good result. But if the image is blurry, warped or noisy, or if the characters are merged, it is likely that the output gets worse than it would otherwise.[1, 2] The output accuracy will also depend on the image resolution.

OCR software is generally better at recognizing machine writing than handwriting. In machine writing the same character always looks the same, given that the same font is used. With handwriting the characters differ a lot depending on who is writing, and even if the same person has written a text, the same character will be a little different each time it is written. Machine writing often consist of more distinct characters, and since handwriting is more varied than machine writing, it is harder for the algorithm to match the characters to its database.

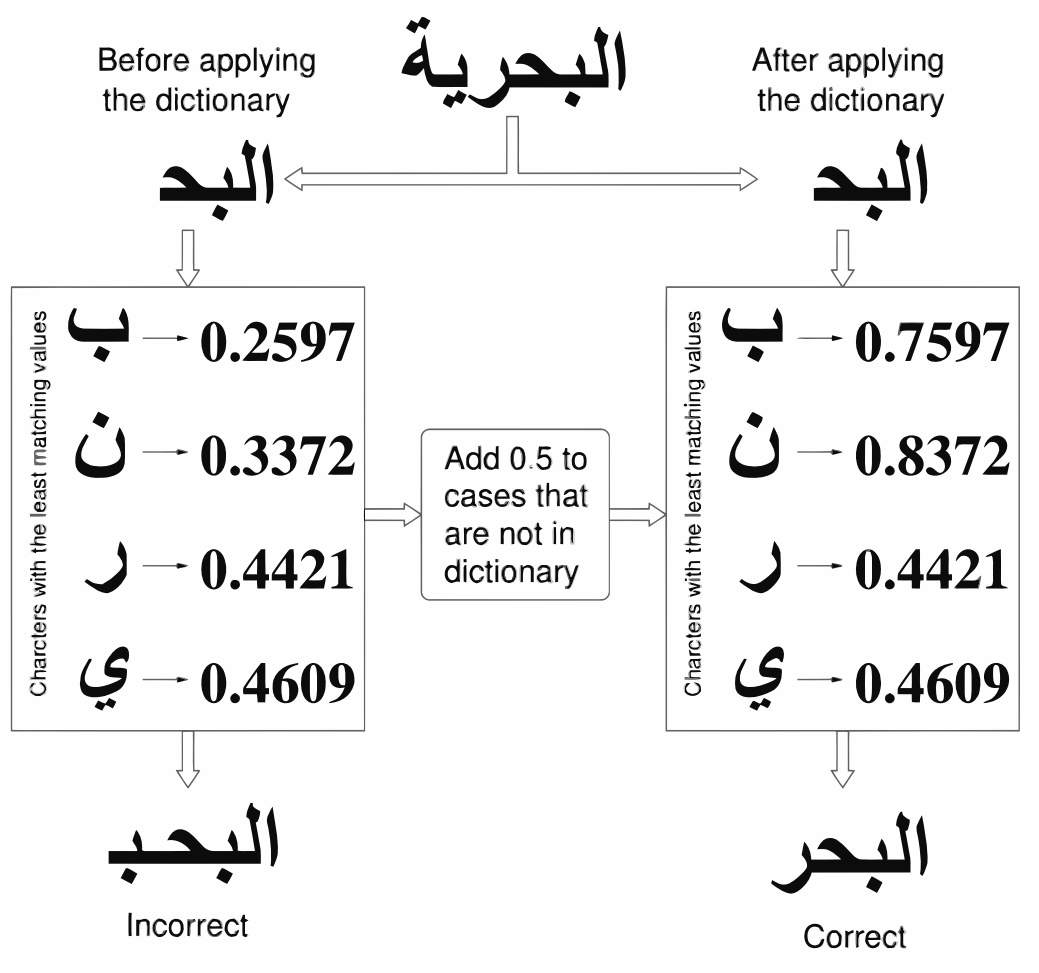


Figure 1: *The correct word is found by using a dictionary. The first letter is the one with the best matching value, but it gets a penalty since it result in a non-dictionary word. With the penalty values added, it is the third letter that has the best matching value, and results in the correct word.[1]*

3 Method

The different software used in the comparison, which images that are included, and how the comparison is done are presented here.

3.1 Software

The software that will be used in the comparison are presented below. Tesseract, Ocrad, CuneiForm, GOCR and OCRopus are open source tools, TOCR,

Abbyy CLI OCR, Leadtools OCR SDK and OCR API Service are commercial tools.

The tools will not be trained for the comparison, and the settings will not be changed depending on the image, the most basic settings will be used in all of the tools to get a fair comparison. If it is possible to specify which language to use in the scans, Swedish or English will be used depending on the image.

Tesseract Tesseract is an open source OCR tool that HP started to develop in 1985. HP continued the development until 1995, and since it was released under the Apache License in 2005 the development has been sponsored by Google. Tesseract can recognize text in over 60 different languages, which have to be downloaded and added to Tesseract manually. If no language is specified for the scan, Tesseract will use English as default, Tesseract can also use multiple languages in each scan. Tesseract supports machine learning, which can be used if the language or font to be scanned doesn't exist.[5]

Ocrad Ocrad is released under the GNU GPL and is an open source OCR tool that has been developed since 2003. It is possible to define which character sets to recognize in the program, which helps narrow the character search. Ocrad can only read three different image formats so the images may need to be converted before the OCR scan.[6]

CuneiForm CuneiForm is an OCR tool that can recognize more than twenty languages and it uses a dictionary to help with the recognition. It is a tool that has been open source under the BSD license since 2008. It is built on a tool from a Russian company named Cognitive Technologies, and the documentation is in Russian. The tool only supports one image format which means that the images may need to be converted before the scan.[7]

GOCR GOCR (sometimes JOCR due to a name conflict) is an open source OCR tool released under the GNU GPL. Images may need to be converted to be used with GOCR since it can only read a few image formats.[8]

OCRopus OCRocpus is an OCR tool released under the Apache License, which means it is an open source tool. OCRopus is sponsored by Google, the Mellon Foundation, and the BMBF TextGrid project, and it is mainly developed at IUPR research group. The recognition can be helped with machine learning, where OCRopus can be trained to recognize new fonts or languages.[9]

TOCR TOCR is a commercial OCR tool that has support for eleven languages, the user does not have to specify which language to use, that is done by TOCR. To help the recognition, TOCR can use character frequency/probability tables, which could help recognize misspellings. It can only read two different image formats, so there may be need for image conversion.[10]

Abbyy CLI OCR Abbyy CLI OCR is a commercial OCR tool that is based on Abbyy FineReader Engine. It supports not only horizontal text but vertical as well, which is specified when running the program. It can read 190 different languages, and offers dictionary support for some of them.[11] Abbyy offers dictionary support for both Swedish and English.

Leadtools OCR SDK Leadtools can recognize over 30 languages, and it uses spell checking and dictionary support to get a more accurate output. It is a commercial OCR tool. It is possible to get an output with the same look as the image, with the same font and layout as the image.[12]

OCR API Service This is a commercial online OCR cloud service, where the image and language is submitted through a HTTP POST request. OCR API Service has support for almost 40 different languages.[13]

3.2 Input images

Different kinds of images will be OCR scanned to see if some software is better at recognizing a specific kind of image, e.g. handwriting or noisy images. The images differ in quality, and some of them are skewed, underlined or contain pictures in the text to see how the software handles that. Most of the images contain text in Swedish, some of them are in English or a combination of Swedish and English. All of the scanned images can be seen in Appendix A.

3.3 Comparison

The comparison of the output from the OCR software to the correct text will be done using the Wagner-Fischer[4] algorithm. The algorithm measures how different two strings are by calculating the edit distance between them. The edit distance is the minimum number of edit operations that is needed for one of the strings to change into the other. The edit operations supported in the algorithm are substituting, deleting or inserting one character in one of the strings.[4] It is possible to assign different costs to the different edit

operations, but in this comparison the cost will be one for each operation. Using one as cost independent of the operation will result in a maximum edit distance equal to the length of the longer string. This corresponds to the case where every character has to be modified by one of the edit operations. If the strings are equal, no character needs to be modified and the edit distance will be zero. Pseudo code of the Wagner-Fischer algorithm with all costs equal to one can be seen in Algorithm 1, and an example of how the edit distance is calculated is shown in Figure 2.

The edit distance will be divided with the length of the longer string (the worst case edit distance), and multiplied by 100, to get a percentage value of the characters that do not match. The percentage output error that will be used for the comparison is defined by

$$\frac{d(s1, s2)}{max\{|s1|, |s2|\}} \cdot 100$$

where d is the edit distance between the strings s1 and s2, and $|s|$ is the length of the string s.

The mean error value will also be calculated for each of the software to get an overall error value that easily can be compared.

Algorithm 1 The Wagner-Fischer algorithm with all costs equal to one[4]

Input: String s1, string s2
Output: Edit distance between s1 and s2

```
D[0..|s1|, 0..|s2|]
for i = 0 to |s1|
    D[i, 0] = i
for j = 0 to |s2|
    D[0, j] = j
for i = 1 to |s1|
    for j = 1 to |s2|
        if s1[i] = s2[j]
            D[i, j] ← D[i − 1, j − 1]
        else
            d1 ← D[i − 1, j − 1] + 1
            d2 ← D[i − 1, j] + 1
            d3 ← D[i, j − 1] + 1
            D[i, j] ← min{d1, d2, d3}
return D[|s1|, |s2|]
```

		a	n	g	e	l	i	c	a
	0	1	2	3	4	5	6	7	8
g	1	1	2	2	3	4	5	6	7
a	2	1	2	3	3	4	5	6	6
b	3	2	2	3	4	4	5	6	7
a	4	3	3	3	4	5	5	6	6
s	5	4	4	4	4	5	6	6	7
i	6	5	5	5	5	5	5	6	7
o	7	6	6	6	6	6	6	6	7

Figure 2: *The edit distance between the strings is the number in the bottom right corner. The edit distance between the strings* "angelica" *and* "gabasio" *is 7.*

4 Results

The result of the OCR scans on all images is shown in Table 1. The table shows the error percentage, which means the more correct the output from the OCR is, the lower the percentage value will be. If the output text matches the correct text perfectly, the error will be 0%.

A graphical illustration of the results is shown in Figure 3. All of the output files can be seen in Appendix B. The input images are of different types, which are shown below:

Type	Image
Picture in the text	k, l
Skewed image	c, e
Handwriting	m, parts on e, f, g
Light	h
Noise	i
Stains	j

It is possible that the text in the output files is correct, but if they contain extra, or missing, whitespace that will result in an error in Algorithm 1, since it compares the strings character by character.

The mean error value, in increasing order, for each of the OCR tools is shown below:

TOCR	8.79%
Leadtools	20.06%
Abbyy	24.53%
OCR API Service	27.81%
OCRopus	31.42%
Tesseract	33.9%
CuneiForm	37.68%
Ocrad	50.25%
GOCR	64.46%

Image (shown in Appendix A)	*Software*				
	Tesseract	Ocrad	CuneiForm	GOCR	OCRopus
a	17.33%	27.45%	15.5%	51.13%	13.31%
b	16.78%	28.96%	23.43%	62.62%	13.86%
c	55.71%	40.7%	19.46%	52.04%	29.25%
d	31.88%	57.87%	49.01%	40.96%	13.5%
e	16.88%	36.2%	17.33%	77.68%	34.77%
f	16.4%	51.51%	17.68%	78.53%	34.34%
g	37.19%	75.78%	7.44%	62.86%	21.86%
h	45.82%	75.78%	7.44%	62.86%	21.86%
i	88.93%	87.72%	84.65%	92.6%	81.91%
j	62.79%	85.73%	8.37%	72.34%	36.36%
k	13.52%	28.19%	85.63%	46.47%	18.91%
l	2.08%	12.5%	35.5%	72.15%	37.96%
m	68.75%	77.59%	77.19%	91.46%	75.51%
n	0.57%	17.56%	78.84%	38.74%	6.49%

Table 1: *Results of the OCR scans on each image from each of the open source tools.*

Image (shown in Appendix A)	*Software*			
	TOCR	Abbyy	Leadtools	OCR API Service
a	1.37%	8.66%	3.4%	8.82%
b	1.6%	7.78%	3.5%	8.23%
c	1.95%	11.51%	8.18%	7.08%
d	1.52%	13.89%	5.01%	20.44%
e	11.89%	16.17%	18.78%	19.3%
f	7.83%	10.82%	23.9%	9.85%
g	3.65%	29.07%	10.3%	53.95%
h	3.65%	29.07%	5.80%	19.92%
i	18.26%	100%	97.35%	100%
j	3.65%	13.92%	28.57%	15.81%
k	3.71%	29.95%	11.95%	19.52%
l	3.19%	7.39%	12.27%	4.8%
m	55.77%	63.27%	51.43%	100%
n	4.99%	1.88%	0.44%	1.66%

Table 1: *Continued. Results of the OCR scans on each image from each of the commercial tools.*

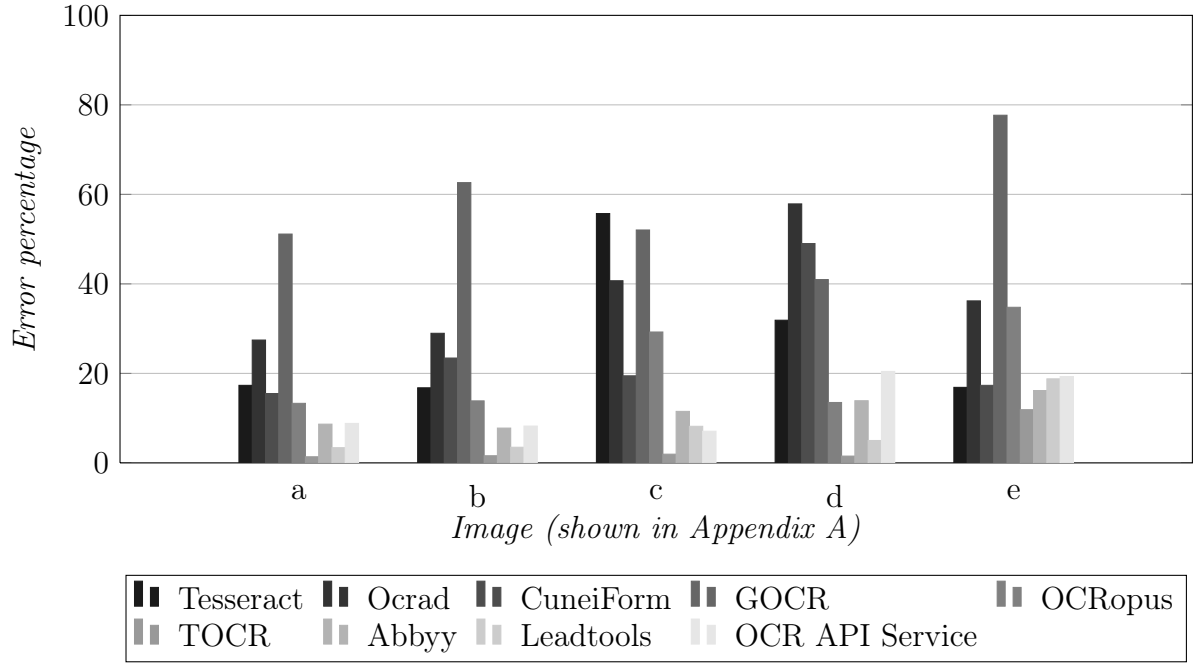


Figure 3: *Illustration of the OCR scan results.*

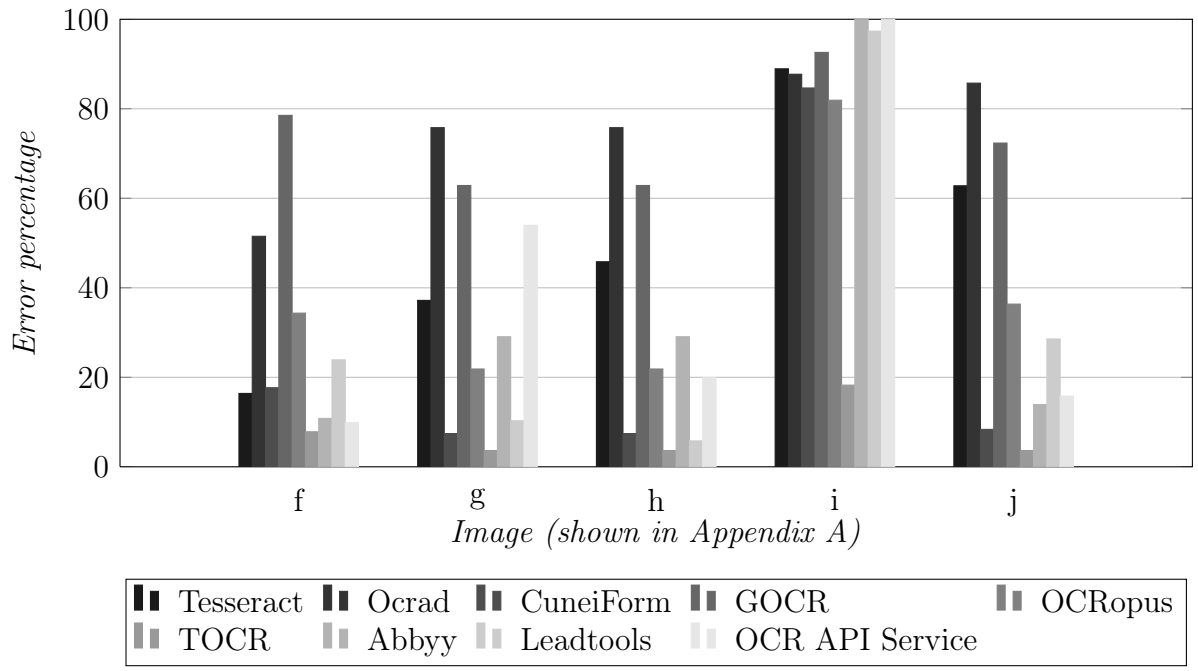


Figure 3: *Continued. Illustration of the OCR scan results.*

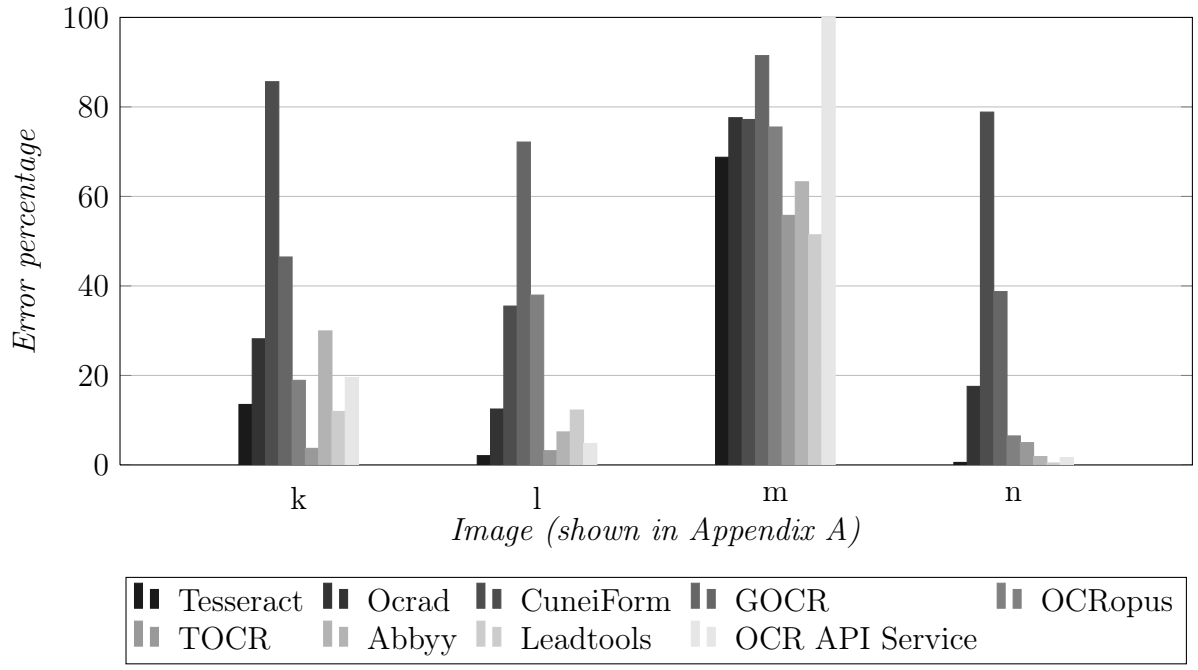


Figure 3: *Continued. Illustration of the OCR scan results.*

5 Discussion

The purpose of the thesis was to test different OCR software on different kinds of images to find out which software generates the most accurate output.

5.1 Software

Some of the commercial tools offer trial versions, limited in time or number of scans. It is difficult to limit the comparison to a certain number of scans, since some testing is needed to understand how the software works, and some of the time limited trial versions did not last long enough for the purpose. This limited the commercial tools that could be used. Emails were sent to the companies to see if they could offer a student version that could be used for the thesis. Some of them offered a version that could be used, and some just pointed at the original trial version.

It was sometimes hard to find documentation or examples on how to use the open source software. The support may also be limited when it comes to open source software, but most of them offer an email address or a forum for that.

5.2 Results

The result of each OCR scan is shown in section 4, the input images and the output files can be seen in Appendix A and B.

Image with picture Image k and image l contains pictures within the text, and the OCR tools that are best at recognizing these images are TOCR and Tesseract. On these images, TOCR has a mean error value of 3.45%, and Tesseract 7.8%.

Both TOCR and Tesseract ignore the picture in image l, and output files 82 (TOCR image l, 3.19%) and 12 (Tesseract image l, 2.08%) do not contain any extra characters where the picture is.

On image k, TOCR recognizes the numbers in the chart, but ignores the rest of it (output file 81, 3.71%). It seems like Tesseract tries to recognize the chart in image k, and output file 11 (Tesseract image k, 13.52%) contains some random characters where the chart is. TOCR also ignores the bars in the legend, Tesseract tries to recognize them and outputs some characters.

Leadtools is more accurate than Tesseract on image k with 11.95% errors compared to Tesseract with 13.52%. The output from these

tools (Leadtools: output file 109, Tesseract: output file 11) does not differ that much, Tesseract outputs more whitespace between the lines and misses most of the decimal points in the image. Since Leadtools recognizes the decimal points, I agree that the output from Leadtools is more accurate than the one from Tesseract on this image.

OCR API Service generates 19.52% errors on image k, and as seen in output file 123 (OCR API Service image k) the text is still very readable. There are some errors in the output, but most of the errors from Algorithm 1 comes from whitespace differences.

Image k contains vertical text as well, and Tesseract is the only tool that recognizes a part of it correctly as seen in output file 11.

CuneiForm and GOCR are the tools that has the highest mean error percentage on image k and l, 60.57% errors for CuneiForm and 59.31% for GOCR. Using one of those would be a bad choice if the images being scanned contain pictures.

Skewed image The skewed images are image c and e, TOCR is the tool with the most accurate result on these images, it produces a mean error percentage of 6.92%. The tool that has the highest mean error on these images is GOCR, with 64.86% errors.

Image c is a skewed version of image b, and most of the OCR tools produce a less accurate output on this image than on the original image. However, CuneiForm, GOCR and OCR API Service produce a better result on the skewed version. It is possible that it is a coincidence that the skewed image gets a lower error value. Output file 31 (CuneiForm image c, 19.46%) contains less line breaks than output file 30 (CuneiForm image b, 23.43%), and some of the lines are only recognized in the skewed case, while some other lines are only recognized in the original case. GOCR produces more whitespace between the letters in the skewed case (output file 45, 52.04%) compared to the original case (output file 44, 62.62%), which makes the text harder to read. The output texts from OCR API Service (skewed: output file 115, 7.08%, original: output file 114, 8.23%) are similar, but some of the errors are only present in one of the output texts.

Handwriting Image e and image f contains some handwritten numbers and image g contains one handwritten character. None of the OCR tools succeeds in recognizing the handwritten parts correctly on these three images. In some cases they get some of the handwritten characters right, but mostly they either ignores the handwriting or they

output what seems to be random characters.

Image m is an image with only handwriting. As seen in section 4 and in Appendix B, the output from the OCR tools is very inaccurate on this image. The software with the most accurate output on image m is Leadtools with 51.43% errors, and as seen in output file 111 the output differ a lot from the correct text.

None of the tested software claims that they can recognize handwritten text, and since all of them produce a very inaccurate output on the handwritten image, a lot of corrections need to be done in the output files. In this case it will probably be better not to use OCR scanning at all.

Noise and stains The images with added noise and stains are image h, i and j, which are all variations of image g. On the original image g, the most accurate tools are TOCR (3.65%), CuneiForm (7.44%) and Leadtools (10.3%).

Most of the OCR tools generate the same amount of errors on the lighter image h as on the original, Tesseract is the only one with more errors, 45.82% compared to 37.19% on the original image. Leadtools and OCR API Service produce better results on the lighter image than on the original. What can be seen in output file 105 (Leadtools image g, 10.3%) and 106 (Leadtools image h, 5.80%) is that the outputted text is almost the same on the two images, but the text layout differ, and the software produces some noise in the end of the text on the original image.

All of the tools are less accurate on the version with more noise, image i, and most of the error percentages differ a lot on the two images. TOCR (18.26%) is the tool with the lowest error on this image, this result is much better than the result from any of the other tools. The output from Leadtools on this image generates 97.35% errors, and output file 107 (Leadtools image i) consists only of a lot of random characters, and is not readable at all. Abbyy and OCR API Service do not recognize any of the characters in this image, and the output is blank (Abbyy: output file 93, OCR API Service: output file 121), which gives an error percentage of 100%. Even though the output on image i from Abbyy and OCR API Service results in a higher error value than Leadtools according to Algorithm 1, my opinion is that it is better to output nothing than to output lots of random characters. In this case Leadtools gets a better result according to the comparison

algorithm since some of the random characters exist in the correct text as well.

Abbyy and OCR API Service are more accurate on the stained version (image *j*) of the image than on the original (image *g*). Abbyy produces 13.92% on the stained version and 29.07% on the original image, OCR API Service produces 15.81% compared to 53.95% on the original. TOCR has the same error percentage as on the original image, while the other tools are less accurate on this image. The tools with the least amount of errors on this image are TOCR with 3.65% errors and CuneiForm with 8.37%.

Image *a* and *b* These two images are of the same document, but on image *b* there is some noise due to the paper behind showing through. On image *a*, there is no paper behind, reducing the amount of noise in the image compared to image *b*. Three of the OCR tools generate a lower error percentage on image *b*, these are Tesseract, Abbyy and OCR API Service, the rest of the tools is more accurate on image *a*. The error percentage on these two images doesn't differ that much, the tools that differ the most are CuneiForm with 15.5% errors on image *a* and 23.43% errors on image *b*, and GOCR with 51.13% errors on image *a* and 62.62% errors on image *b*.

The output from CuneiForm on image *b* (output file 30, 23.43%) misses some lines or parts of some lines in the image, which is recognized on image *a* (output file 29, 15.5%). CuneiForm recognizes "mZ" instead of "m2" in some places on image *b*, but on image *a*, the correct text "m2" is recognized. There are some other differences in the output as well, and because of the differences above, I agree that the output on image *a* is better than the one on image *b*.

GOCR is inaccurate on both of these images, and it is hard to decide if the output on image *a* (output file 43, 51.13%) really is better and easier to read than the output on image *b* (output file 44, 62.62%). Even if the result is better on image *a*, it is still more than 50% errors in the output.

Image *n* This image contains only text and minimal noise, and many of the tools produce a good and readable output on this image. Leadtools is the most accurate tool with 0.44% errors, followed by Tesseract with 0.57% errors. CuneiForm is the tool with the highest error percentage, 78.84%, and output file 42 (CuneiForm image *n*) is not readable at all and only contains random characters.

Abbyy Abbyy produces a better output on image j (13.92%) than image g (29.07%), which is a bit odd, since image j is a stained version of image g. It is possible that it is just a coincidence that the error percentage is lower, but output file 94 (Abbyy image j) does contain less noise and less whitespace than output file 91 (Abbyy image g). Another possibility to why the output is more accurate on this image may be that the software uses a different noise threshold on images with a lot of noise, and that it is able to remove more of the disturbances in the images with more noise, which results in a more accurate output.

TOCR TOCR generates the most accurate result on most of the images in the comparison. Tesseract is the only software that is more accurate on image l, where Tesseract produces an output containing 2.08% errors, and TOCR produces 3.19% errors. The output file from TOCR on this image (output file 82) contains the correct text, but it misses whitespace between the lines. Tesseract doesn't recognize the page number on this image (output file 12), but it recognizes the whitespace, which results in Tesseract getting a lower error percentage than TOCR on this image.

On image n, TOCR generates 4.99% errors, and output file 84 shows that TOCR fails to recognize the letter "d", it recognizes "cl" instead. Apart from this, the output text seems very accurate.

Since TOCR only reads two different image formats, a limitation is that images of other formats need to be converted before the scan. It might be worth converting the images and use TOCR, since it produces much more accurate results than any of the other tools in this comparison.

OCR API Service As seen in output files 121 (OCR API Service image i) and 125 (OCR API Service image m), OCR API Service does not produce any output on images i and m. Image i is the image with added noise, and image m is the handwritten image.

OCR API Service generates a more accurate result on the lighter image h (output file 120, 19.92%) and the stained image j (output file 122, 15.81%) than on the original image g (output file 119, 53.95%). It seems like the tool ignores the right side of the image in the original case (output file 119), and therefore misses some of the characters in the image. Since this issue does not appear in the lighter and in the stained version of the image, those results are more accurate.

OCRopus The most accurate open source software in this comparison is OCRopus, with a mean error of 31.42%. The tool is not the most

accurate on any of the images in Appendix A, but it is the most accurate open source tool on some of the scanned images.

The most accurate output from OCRopus is 6.49% errors on image n. Output file 70 (OCRopus image n) contains some recognition errors, but it is still possible to understand most of the output text.

On image i, OCRopus generates 81.91% errors, which is the most inaccurate result from the tool. This is the image with added noise, and OCRopus is the second most accurate software on this image, TOCR is the only tool that is more accurate with 18.26% errors. Output file 65 (OCRopus image i) contains only random characters, and the output text does not match the text in the image at all.

OCRopus is very slow, much slower than any of the other tools, and even though it is the most accurate open source software in the comparison, I would rather use another tool if I wanted an open source software.

Tesseract Tesseract is the second most accurate open source tool in this comparison, it produces a mean error of 33.9%. It is the most accurate tool on image l with 2.08% errors. On image n, Tesseract generates 0.57% errors, and Leadtools is the only tool with a more accurate output on this image with 0.44% errors. The text in output files 12 (Tesseract image l) and 14 (Tesseract image n) contains some recognition errors, but the text is very readable.

Tesseract is the least accurate tool on image c, which is one of the skewed images, with 55.71% errors (output file 3). The tool recognizes some parts of the image correctly, but it misses many of the lines and the output contains some noise as well.

GOCR This is the least accurate software in this comparison, with a mean error of 64.46%.

GOCR is not the least accurate software on all of the images, but it is among the least accurate tools on every image. The output files from GOCR (section B.4) contain a lot of noise, and are hard to read. The most accurate output from GOCR is 38.74% on image n, and as seen in output file 56 (GOCR image n), the text contains a lot of errors and noise and is not readable at all.

The other open source tools tested in this comparison are better choices than GOCR for OCR scanning, at least on images like the ones in this comparison.

5.3 Open source vs. commercial software

The mean error results from the comparison shows that the commercial tools are more accurate than any of the open source tools. On some images there is an open source tool that produces a lower error value, or close to the value of the commercial tools.

The most accurate open source tool is OCRopus, which produces a mean error of 31.42%, compared to the least accurate commercial tool OCR API Service with 27.81% errors. Tesseract is the open source tool closest to the results of OCRopus, with a mean error of 33.9%, and since the mean results does not differ that much and OCRopus was very slow in the tests, Tesseract would probably be a better choice than OCRopus.

The tools with the highest mean error in the comparison are Ocrad with 50.25% errors and GOCR with 64.46%. The output from these tools is very inaccurate and I think that none of the output files in Appendix B.2 (Ocrad) and B.4 (GOCR) contains an acceptable result. The most accurate result from these tools is from Ocrad on image l with 12.5% errors, and there are a lot of errors in output file 26 (Ocrad image l).

TOCR is the software with the most accurate output (mean error 8.79%), the tool that is closest to this result is Leadtools with a mean error of 20.06%. The mean error percentages from these tools are not that close, and TOCR generates a more accurate result than Leadtools on most of the images in Appendix A, Leadtools is more accurate than TOCR on two images, image m and n. The error percentages from TOCR and Leadtools on image m (TOCR: 55.77%, Leadtools: 51.43%) does not differ that much, and the output from both of the tools is inaccurate and very hard to read (TOCR: output file 83, Leadtools: output file 111). On image n, TOCR generates 4.99% errors and Leadtools 0.44%, this is the image where TOCR fails to recognize the ”d”s in the text (output file 84). Since Leadtools (output file 112) does not have this problem, that output is more accurate than the one from TOCR.

6 Conclusion

As seen in section 4, the OCR tool with the lowest mean error value on all of the images in Appendix A, is the commercial software TOCR with 8.79% errors. The software closest to the results of TOCR is the commercial tool Leadtools, with 20.06% errors. The most accurate open source software is OCRopus with a mean error value of 31.42%, followed by Tesseract with 33.9% errors.

The four commercial OCR tools are the ones that have the lowest mean error values, this shows that it probably would be a good idea to invest in a commercial OCR software, at least if the images being scanned are of different quality, as the ones in this thesis. The results from TOCR and the other tools differ a lot and the better choice would be TOCR.

If a commercial tool is unwanted, the mean error shows that OCRopus is the most accurate open source tool, but since it is very slow, Tesseract, which is the second most accurate open source tool, would be a better choice, at least if the scanning time is an issue. The mean error does not differ that much between these tools, OCRopus generates a mean error of 31.42% and Tesseract 33.9%.

Ocrad and GOCR are the software that has the highest mean error values (Ocrad: 50.25%, GOCR: 64.46%) in this comparison, using one of these tools would be a bad choice if the images that is going to be OCR scanned are similar to the ones in Appendix A.

References

[1] Inad Aljarrah, Osama Al-Khaleel, Khaldoon Mhaidat, Mu'ath Alrefai, Abdullah Alzu'bi, Mohammad Rabab'ah, *Automated System for Arabic Optical Character Recognition with Lookup Dictionary*, Journal of Emerging Technologies in Web Intelligence, Nov 2012, Vol. 4 Issue 4, pp. 362-370

[2] Tobias Blanke, Michael Bryant, Mark Hedges, *Open source optical character recognition for historical research*, Journal of Documentation, Vol. 68 Iss: 5, pp. 659-683

[3] Herbert F. Schantz, *The history of OCR: Optical character recognition*, Recognition Technologies Users Association, 1982

[4] Robert A. Wagner, Michael J. Fischer, *The String-to-String Correction Problem*, Journal of the Association for Computing Machinery, Vol. 21, No. 1, January 1974, pp. 168-173

Software used

[5] Tesseract-ocr, `http://code.google.com/p/tesseract-ocr/`

[6] Ocrad, `http://www.gnu.org/software/ocrad/`

[7] CuneiForm, `http://cognitiveforms.com/ru.html#1189-CuneiForm`

[8] GOCR, `http://jocr.sourceforge.net/`

[9] OCRopus, `https://code.google.com/p/ocropus/`

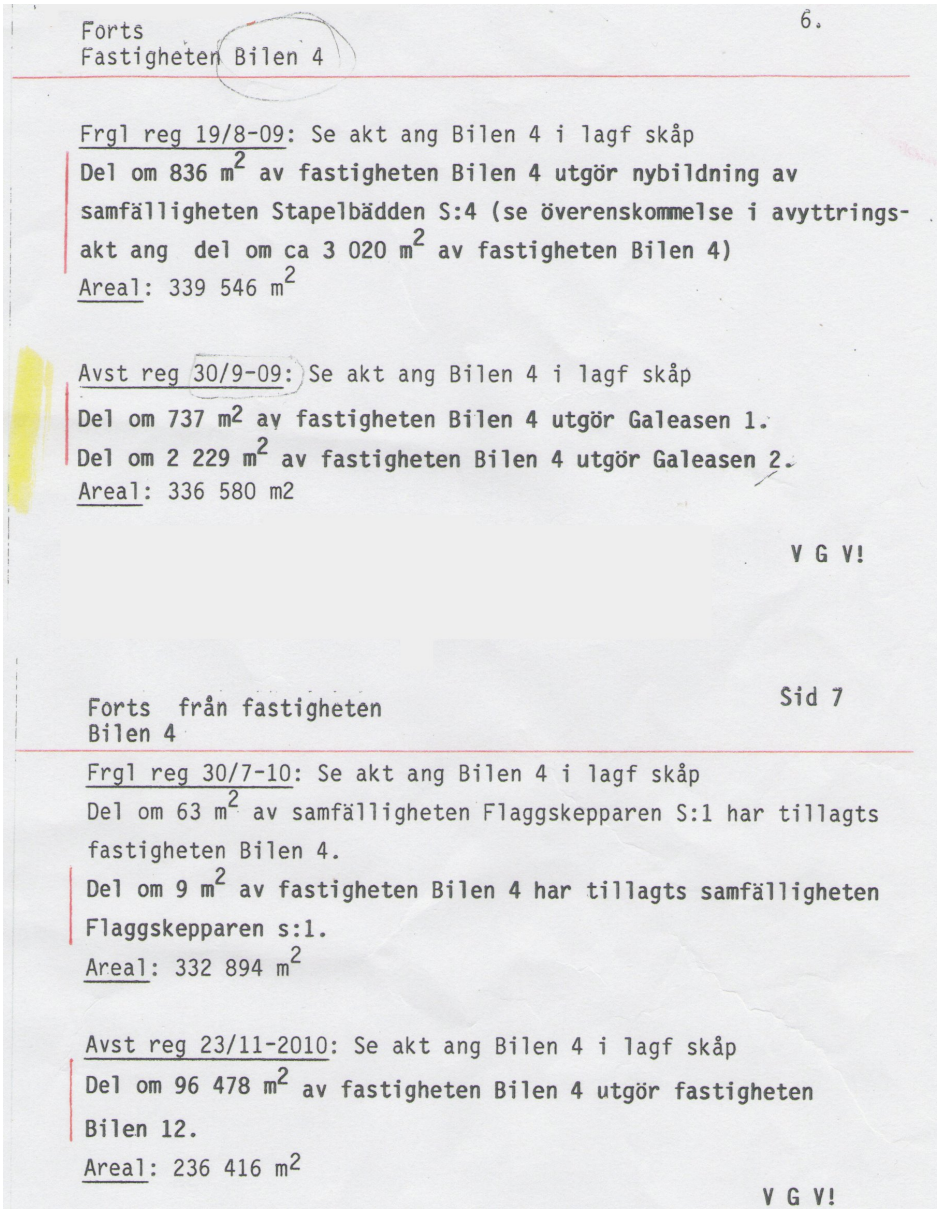
[10] TOCR, `http://www.transym.com/tocr-the-integrators-choice.htm/`

[11] Abbyy CLI OCR, `http://www.ocr4linux.com/`

[12] Leadtools OCR SDK, `http://www.leadtools.com/sdk/ocr/`

[13] OCR API Service, `http://ocrapiservice.com/`

Appendix A: Input images



Forts
Fastigheten Bilen 4

6.

Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m^2 av fastigheten Bilen 4 utgör nybildning av samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-akt ang del om ca 3 020 m^2 av fastigheten Bilen 4)
Areal: 339 546 m^2

Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m^2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m^2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2

V G V!

Forts från fastigheten
Bilen 4

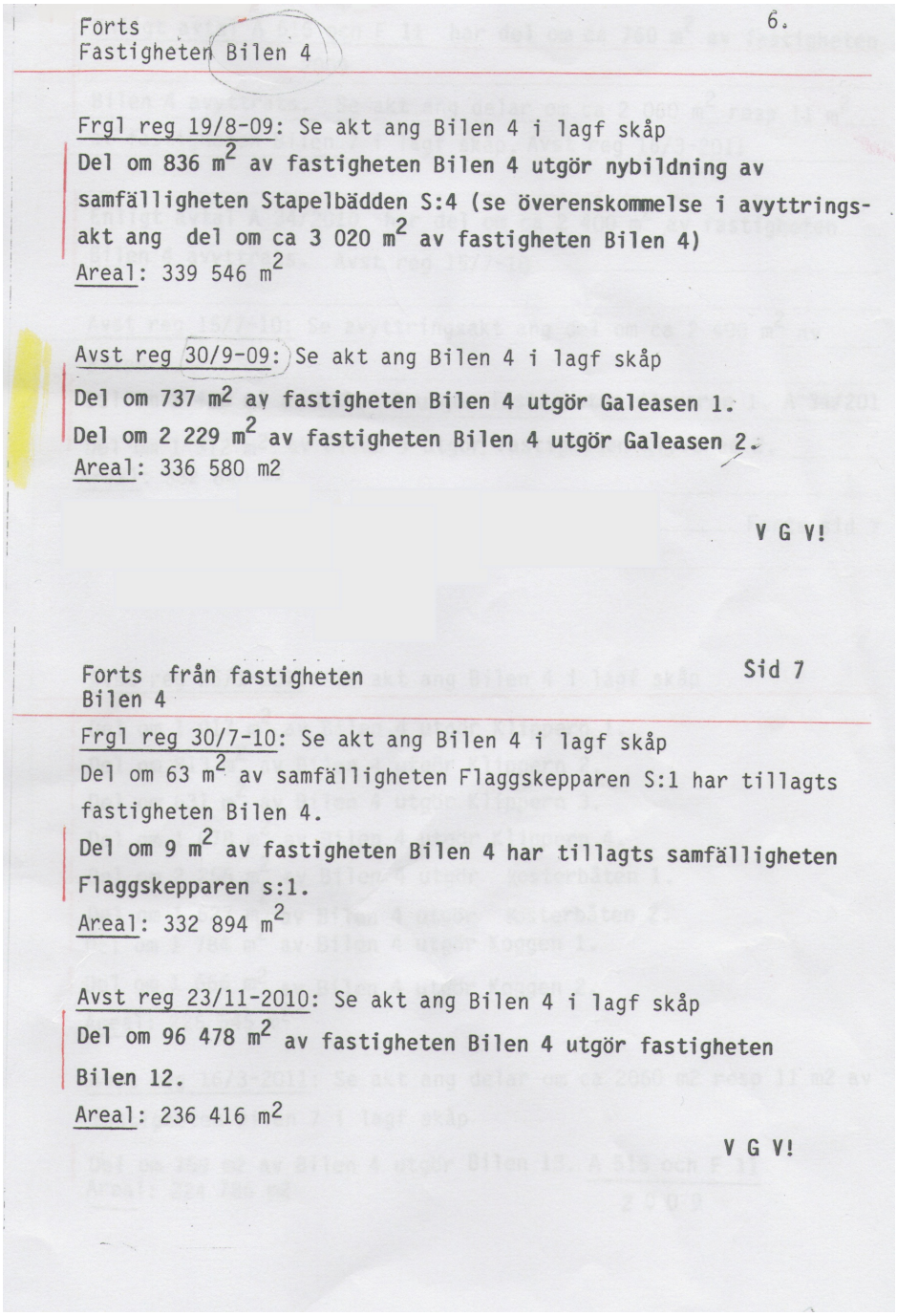
Sid 7

Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m^2 av samfälligheten Flaggskepparen S:1 har tillagts fastigheten Bilen 4.
Del om 9 m^2 av fastigheten Bilen 4 har tillagts samfälligheten Flaggskepparen s:1.
Areal: 332 894 m^2

Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m^2 av fastigheten Bilen 4 utgör fastigheten Bilen 12.
Areal: 236 416 m^2

V G V!

Image *a*



Forts 6.
Fastigheten Bilen 4

Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m^2 av fastigheten Bilen 4 utgör nybildning av samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-akt ang del om ca 3 020 m^2 av fastigheten Bilen 4)
Areal: 339 546 m^2

Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m^2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m^2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2

V G V!

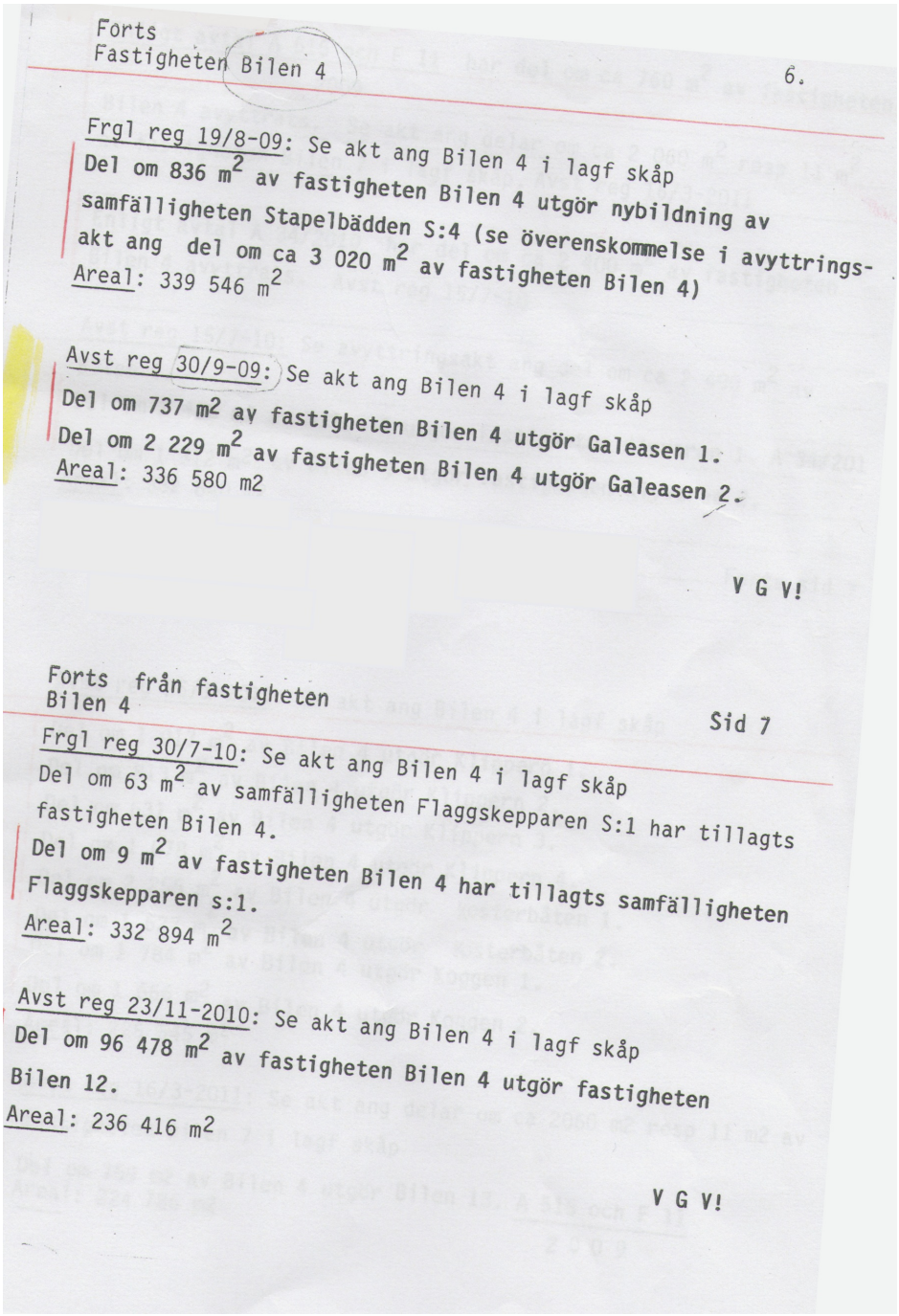
Forts från fastigheten Sid 7
Bilen 4

Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m^2 av samfälligheten Flaggskepparen S:1 har tillagts fastigheten Bilen 4.
Del om 9 m^2 av fastigheten Bilen 4 har tillagts samfälligheten Flaggskepparen s:1.
Areal: 332 894 m^2

Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m^2 av fastigheten Bilen 4 utgör fastigheten Bilen 12.
Areal: 236 416 m^2

V G V!

Image *b*: *Almost the same as image* a, *but some text from another page is visible through the paper.*



Forts
Fastigheten Bilen 4 6.

Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m^2 av fastigheten Bilen 4 utgör nybildning av samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-akt ang del om ca 3 020 m^2 av fastigheten Bilen 4)
Areal: 339 546 m^2

Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m^2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m^2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2

V G V!

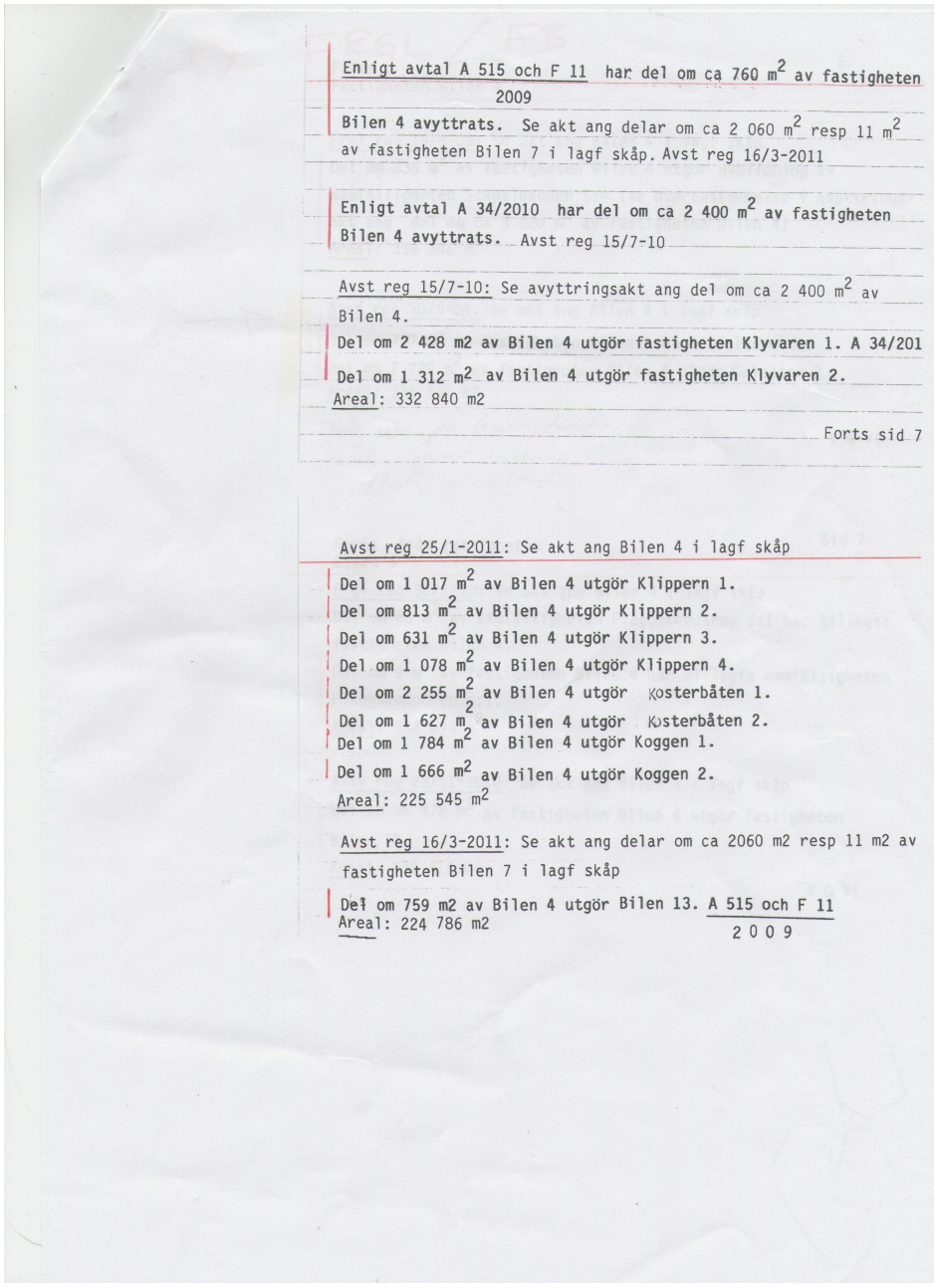
Forts från fastigheten
Bilen 4 Sid 7

Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m^2 av samfälligheten Flaggskepparen S:1 har tillagts fastigheten Bilen 4.
Del om 9 m^2 av fastigheten Bilen 4 har tillagts samfälligheten Flaggskepparen s:1.
Areal: 332 894 m^2

Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m^2 av fastigheten Bilen 4 utgör fastigheten Bilen 12.
Areal: 236 416 m^2

V G V!

Image *c*: *A skewed version of image* b.



Enligt avtal A 515 och F 11 har del om ca 760 m^2 av fastigheten
2009
Bilen 4 avyttrats. Se akt ang delar om ca 2 060 m^2 resp 11 m^2
av fastigheten Bilen 7 i lagf skåp. Avst reg 16/3-2011

Enligt avtal A 34/2010 har del om ca 2 400 m^2 av fastigheten
Bilen 4 avyttrats. Avst reg 15/7-10

Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m^2 av
Bilen 4.
Del om 2 428 m2 av Bilen 4 utgör fastigheten Klyvaren 1. A 34/201
Del om 1 312 m^2 av Bilen 4 utgör fastigheten Klyvaren 2.
Areal: 332 840 m2

Forts sid 7

Avst reg 25/1-2011: Se akt ang Bilen 4 i lagf skåp

Del om 1 017 m^2 av Bilen 4 utgör Klippern 1.
Del om 813 m^2 av Bilen 4 utgör Klippern 2.
Del om 631 m^2 av Bilen 4 utgör Klippern 3.
Del om 1 078 m^2 av Bilen 4 utgör Klippern 4.
Del om 2 255 m^2 av Bilen 4 utgör Kosterbåten 1.
Del om 1 627 m^2 av Bilen 4 utgör Kosterbåten 2.
Del om 1 784 m^2 av Bilen 4 utgör Koggen 1.
Del om 1 666 m^2 av Bilen 4 utgör Koggen 2.
Areal: 225 545 m^2

Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skåp

Del om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 11
2009
Areal: 224 786 m2

Image *d*

Fastigheten Oxie 1:5

Frgl reg 19/4-79: se akt ang fastigh Oxie 1:1 m fl
Del om 10 801,7 m² av samfälligheten Oxie s:2 har tillagts fastigheten Oxie 1:5 (F 29/1979 A 27/1979)

2 234 580,5 m²

Enl. avtal A61, F60 1979 har del av fastigheten Oxie 1:5 avyttrats, se akt ang. ca 78.000 m² av fastigheten Oxie 21;2. Delen har vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 m².

Areal: 2 227 292,6 m²
Enl. avtal A 20/1981 har del om 32.615,9 m² av Oxie 1:5, avs. att bilda fastigheten Fornlämningen 1 avyttrats. (avst reg 9/10-81)
Avst reg 17/3-88: se akt ang Oxie 1:1 m fl
Del om 14 225 m² av Oxie 1:5 utgör Fornlämningen 2
Areal: 2 180 452 m²

Image *e*

657,654

Fastigheten Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1, Oxie 17:2, Oxie 19:1, Oxie 20:8, Oxie 25:9, Oxie 28:1 och Oxie 29:1.
Areal efter sammanläggning: 2.231.222 m²
Särskilt namn enl. reg. beslut den 15/4-92: Fredriksberg
Frgl. reg.17/10-75 : se akt ang. stg 2440 m.fl. i Fosie
Delar om 85,0 m² av samfällda området adj, 1.956,0 m² av samfällda området Oxie s:1, 33.507,2 m² av stadsägan 2834 och 22.122,4 m² av stadsägan 2467, 2468 har tillagts fastigheten Oxie 1:5.

Image *f*

Forts från
Fastigheten Oxie 1:5 4.

Del om 1 261 m^2 av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m^2

Enligt avtal A 329/2006 har del om ca 23 300 m^2 av fastigheten
Oxie 1:5 avyttrats. 1

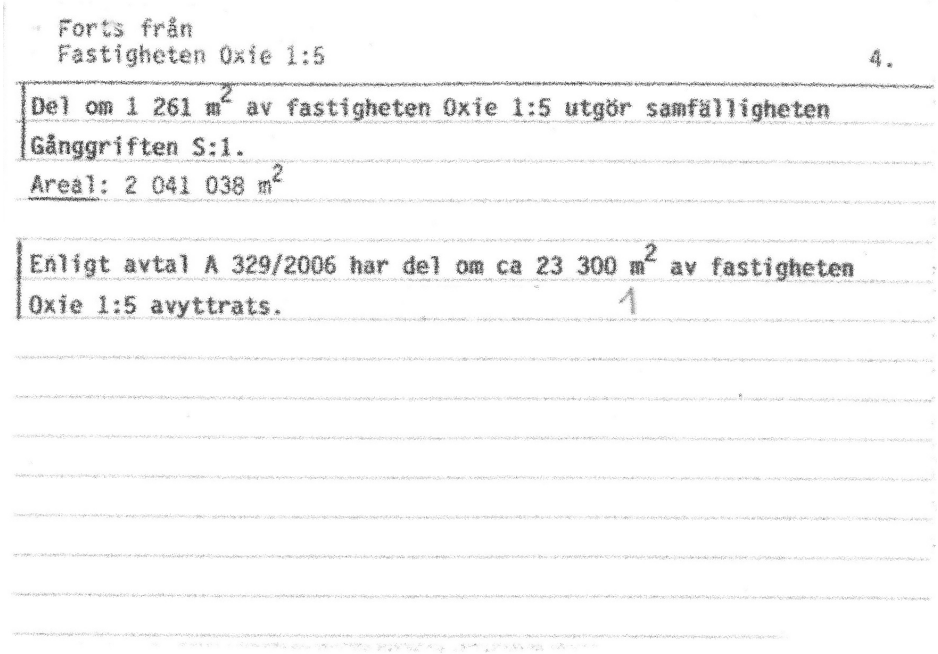
Image *g*

Forts från
Fastigheten Oxie 1:5 4.

Del om 1 261 m^2 av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m^2

Enligt avtal A 329/2006 har del om ca 23 300 m^2 av fastigheten
Oxie 1:5 avyttrats. 1

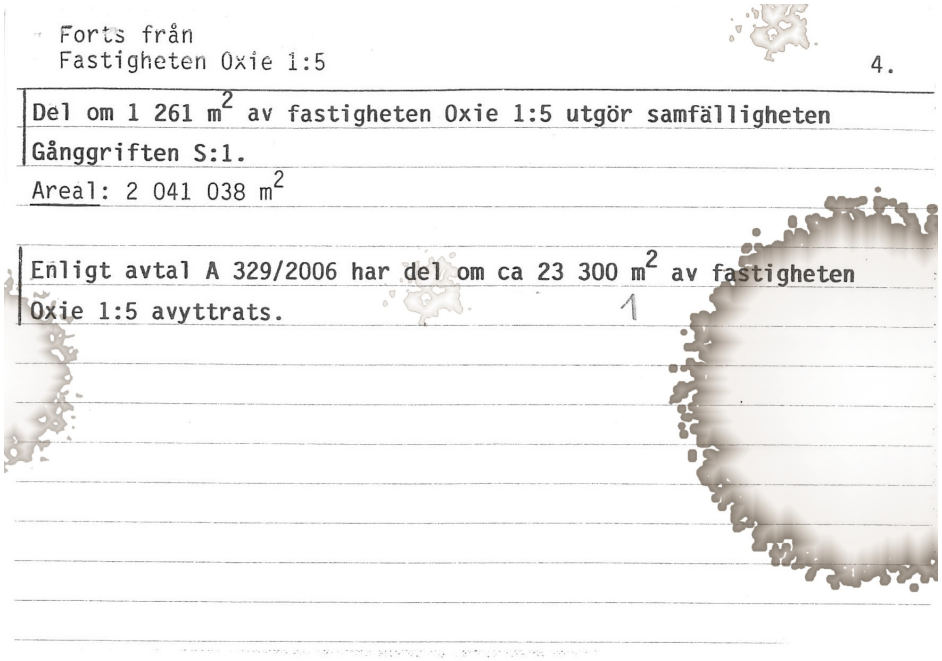
Image *h*: *A lighter version of image* g.


Forts från
Fastigheten Oxie 1:5 4.

Del om 1 261 m^2 av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m^2

Enligt avtal A 329/2006 har del om ca 23 300 m^2 av fastigheten
Oxie 1:5 avyttrats. 1

Image *i*: *A noisier version of image* g.


Forts från
Fastigheten Oxie 1:5 4.

Del om 1 261 m^2 av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m^2

Enligt avtal A 329/2006 har del om ca 23 300 m^2 av fastigheten
Oxie 1:5 avyttrats. 1

Image *j*: *A stained version of image* g.

Software	*Image (shown in Appendix A)*				
	a	b	c	d	e
Tesseract	16.97%	35.59%	16.72%	36.78%	16.4%
Ocrad	28.96%	57.93%	36.2%	75.78%	51.51%
OpenOCR	23.43%	49.01%	17.33%	7.442%	17.68%
GOCR	62.62%	41.04%	77.68%	62.86%	78.53%
OCRopus	32.87%	15.79%	37.64%	41.52%	39.38%
TOCR	4.74%	4.413%	13.22%	7.207%	11.26%
Abbyy	9.513%	29.16%	24.56%	27.03%	15.31%

Table 1: *Results of the OCR scans.*

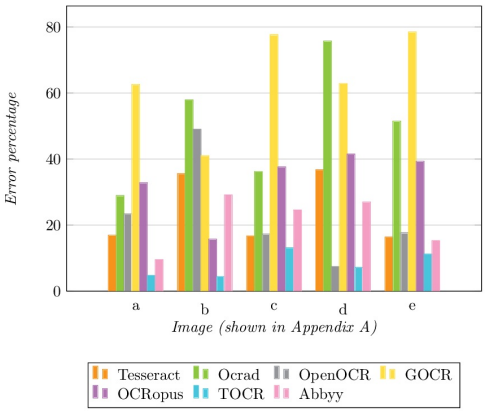


Figure 2: *Illustration of the scan results.*

The mean error value in increasing order for each of the OCR tools is shown below:

TOCR 8.169%
Abbyy 21.11%
OpenOCR 22.98%
Tesseract 24.49%
OCRopus 33.44%
Ocrad 50.07%
GOCR 64.55%

Image k

Testing OCR tools

Angelica Gabasio

1 English

This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.

Figure 1: This is a cat.

2 Swedish

Och så lite svenska tecken. Båten åker i sjön.
Det är blött.

Image l

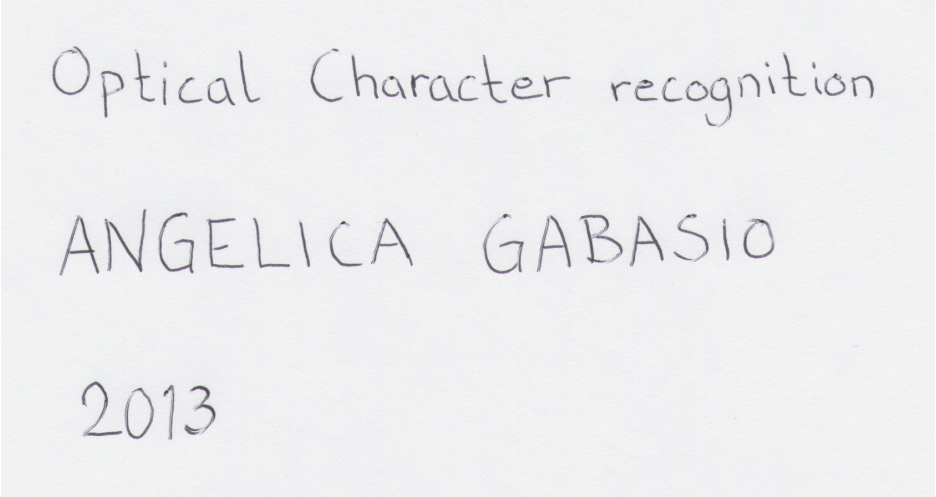


Image m

The algorithm uses a database of letters and numerals to match the characters. The algorithm will calculate a matching value for each character and selects the option with the best value.[1]

A way of improving the output is to check if the scanned word exists in a dictionary. If it doesn't, it is likely that some of the analyzed letters are wrong, and some of the letters may be changed so the word matches a word in the dictionary. This is done by adding a penalty value to the matching value if the analyzed character result in a word not in the dictionary. This makes it possible for another letter to get a better matching value if the word with the new letter is part of the dictionary. The letters are not always changed, sometimes a non-dictionary word gets a better matching score, even with the penalty value, than the one with the changed letters that is in the dictionary. Even if using a dictionary may help the output accuracy it is no guarantee that the correct word will be found.[1] An example of the use of a dictionary can be seen in Figure 1.

Other ways to get a better output is to use machine learning, where the OCR tool can be trained to recognize more languages or fonts, or to use the knowledge of which character combinations are more likely to appear.

The better the image is, the more accurate the result gets. If the OCR algorithm is applied on a horizontal text with very little or no noise and the letters are well separated from each other it is possible to get a very good result. But if the image is blurry, warped or noisy, or if the characters are merged, it is likely that the output gets worse than it would otherwise.[1, 2] The output accuracy will also depend on the image resolution.

OCR software is generally better at recognizing machine writing than handwriting. In machine writing the same character always looks the same, given that the same font is used. With handwriting the characters differ a lot depending on who is writing, and even if the same person has written a text, the same character will be a little different each time it is written. Machine writing often consist of more distinct characters, and since handwriting is more varied than machine writing, it is harder for the algorithm to match the characters to its database.

Image n

Appendix B: Output files

The output from the OCR tools on the images in Appendix A is shown here.

Some of the output files contain rows that are too wide to fit on the page, these files have been forced to break those lines. The original output files, with no extra line breaks, are the ones used in the comparison.

B.1 Tesseract

```
Forts a-   6a

FastigheteáfBi1en 4 *\

Frgi reg 19/8-09: Se akt ang Biien 4 i 1agf skåp

Del om 836 m2 av fastigheten Bilen 4 utgör nybiidning av

samfä11igheten Stapeibädden S:4 (se överenskomme1se i avyttrings- .

2

Ä akt ang de] om ca 3 020 m av fastigheten Bilen 4)

ä Areaiz 339 546 m2



Avst reg  0 /9-09:;se akt ang Biien 4 i 1agf skåp
5 De1 om 737 m2 av fastigheten Biien 4 utgör Gaieasen 1;

De] om 2 229 m2 av fastigheten Bi1en 4 utgör Ga1easen g-
Area1: 336 580 m2 '

V G V!

i
I
1
1

Forts från fastigheten
2 Bi1en 41 ___ __w___m"_m____ _____ VWm_m_____ww
V Frgl reg 30/7-10: Se akt ang Bi1en 4 i 1agf skåp

De] om 63 mz av samfäiiigheten F1aggskepparen 5:1 har tiliagts
fastigheten Bi1en 4.

2 av fastigheten Bi1en 4 har ti11agts samfä11igheten
Fiaggskepparen s:1.
Areai: 332 894 m2

Del om 9 m



Avst reg 23/11-2010: Se akt ang Biien 4 i 1agf skåp

D61 Om 95 478 m2 av fastigheten Bilen 4 utgör fastigheten
Biien 12.

Areal: 236 416 m2



V G V!
```

Output file 1: *Tesseract – Image* a

å
va'

r a' 'my' á\\
r

Forts A, 2 x
n. Bilen 4

, \
Fastighete

f' .\
l

Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av

samfälligheten Stapelbädden S24 (se överenskommelse i avyttrings- .

akt ang del om ca 3 020 m2

Areal: 339 546 m2

av fastigheten Bilen 4)

Avst reg 0 /9-09i;Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1;

Del om 2 229 m2
Areal: 336 580 m2

av fastigheten Bilen 4 utgör Galeasen g-

/

V G V!

Forts från fastigheten Sid 7
Bilen 4- __m" _ _

J

Frgl reg 3073-10: Se akt ang Bilen 4 i l a g f sk åp_m
Del om 63 mz av samfälligheten Flaggskepparen 5:1 har tillagts
fastigheten Bilen 4.

Del om 9 m2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:1.

Areal: 332 894 m2

Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
D61 Om 95 478 M2 av fastigheten Bilen 4 utgör fastigheten

Bilen 12.
Areal: 236 416 m2

```
V G V!
```

Output file 2: *Tesseract – Image* b

```
Forts _ ,  '          '%    \;
Fastigheten/Bilen 4   l

   6.
__1_____________.g___________m- -

Frgl reg 19/8-09: Se akt ang Bilen 4 i l

Del om 836 m2



V G V'
Forts från fastigheten $'d 7
____Å'l1i_ _MM_ .___..M__..._.--__.-.-___,--_..._Me.-._.-_____.____.
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf sk
Del om 63 m2 av samfälli

fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:1.

Areal: 332 894 m2

I Avst reg 23/11-2010:

Del Om 95 478 M2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.

Areal: 236 416 m?
```

Output file 3: *Tesseract – Image* c

```
. ?Eli F_ Vta] A 515  Ch  F 11 har de? om ca 760 m2 av fastighgten

_ ,. ?009 __ W ._
_ 3116"êgvxttratsaz--S9.akatansadeiaráaom ca 2 060 m2 resp li mf

av fa5ti9" e t e   3ÅJ?U-7 _T lâ9f.âkåp-wÅY$t_f69.l6/3'2Qll



gu .Enligt avta1 A 34/2010 ihar de? om ca 05400 m2 av fasttg eten
 d, Bil?" 4 .avyttrats.__ Aysj;  r e g   15/7710_ __ _ _m_ _____._ _

:gÅJstHrEg55AIá77-1qgg5 Se 5$   t t r '1"ägsá L  é  ang 221 omu ca? 460 in? av
0 Bilen_4,     "'_ä__

I 0e1_om z_42e mz ;Q åilep 4 utgår fastjghetéa klyváké   1. Ä 34/261

De1.om.1.31 z m2_av  Bilen_4.utgörwfastighetgn Klyvaren 2:
_ Area]:_332g840 m2_

*vn*--M -  --    -- _u -  - HM --N--4h___4wmuwm__4t_____m_-Eorts sidM7

.I De1 om 1 017 m2 av Bilen 4 utgör Klippern 1.
```

```
..J Del om 813 m2 av Bilen 4 utgör Klippern 2.

{ De] om 631 m2 av Bilen 4 utgör Klippern 3.
I Del om 1 078 m2 av Bilen 4 utgör Klippern 4.

I Del om 2 255 m2 av Bilen 4 utgör Kosterbåten 1.

2
f Del Om 1 627 nä-av Bilen 4 utgör Josterbåten 2.
rDel om 1 784 m av Bilen 4 utgör Koggen 1.

.HI D91  m  1 555 m2 av Bilen 4 utgör Koggen 2.
Areal: 225 545 m2

Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
áfastighéten Bilen 7 i lagf skåp
| 0%: om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 11



' Areal: 224 786 m2 2 0 0 9

4...
```

Output file 4: *Tesseract – Image* d

```
Fastigheten Oxie 1:5

Frgl reg 1214-79: se akt ang fastigh Oxie 1:1 m fl

Del om 10 801,7 Å av samfälligheten Oxie s:2 har tillagts
fastigheten Oxie 1:5 (E â3%13;g)
Z.â3V.5?O,5'07

A61 F60
Enl- aVtal   '"J"'   ' har del av fastigheten Oxie 1:5 avyttrats,e

1979
se akt ang. ca 78.000 Å av fastigheten Oxie 21;2. Delen har

vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 E.

Areal: 2 227 292,6 f
Enl. avtal A 20/1981 har del om 52.615,9 Å av Oxie 1:5, avs.

(MJ M7 ?D 'KID



att bilda fastigheten Fornlämningen 1 avyttrats.
Avst 19% 17/3-88: se akt ang Oxie 1:1 m fl
Del om lü 225 m2 aå Oxie 1:5 utgör Fornlämningen 2
Axsai: 2 180 u52 m
```

Output file 5: *Tesseract – Image* e

```
Q57, Q V
Fastigheten Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1,
```

```
Oxie 17:2, Oxie 19:1, Oxie 20:8, Oxie 25:9, Oxie 28:1 Och
Oxie 2931.

;Area1'effer sam@an1ä"  nin   : 2.231.222 å


Sä  skiltinamn  -eni; reg  beslut den 15/4f921.EL dIÅk D LQ1

Frgl, reg.17{10_Z  : se akt ang. stg 2440 m.fl. i Fosie

Delar om 85,0 J av samfällda området adj, 1.956,0 f av sam-
fällde områäet Oxie s:1. 35.507,2 ä av stadsägan.2834 och

22.12?,4 Å av seadsägan 9á67,2468 har tilla ts fastigheten
```

Output file 6: *Tesseract – Image* f

```
W Forts från
Fastigheten Oxie 1:5 4.

-D61 9m 1?261 iäwav faSti9heten,0 ieh1 =5 utgör  mf ä11i9h ten_-
_Eânüfi fteg! 5:1-,
mArea1: 2 041 038 m2

åå1igtZåvta1 Ä 329[gbÖê:Har;gê1ZgåWcg_23 300 m? gyqfa tigheten
"Oxie 1:5 avyttra;  ;, 4
```

Output file 7: *Tesseract – Image* g

```
W Forts från

Fastigheten Oxie 1:5 4.


_ De1 pm 11261 m_"ay fastjghe eg_O je 1:5 utgör  mf ä11jghçtegn<
_Qânggtiftegy 5:1. _ _ _{  ,i  ,
mArea1: 2 041 038 m2

wiå kligtw åvtaJjÄ ;zái/:z öhdêiwrrarf _áêl Garcia: 2; swmzwçfvqaiastigwsn-
"Oxie 1:5 avyttra;  ;, 4
```

Output file 8: *Tesseract – Image* h

```
?mig ?råa V V A
?vâääääääüäêiâ üxêü  å.


wvxer-w - snw  ,muwwmwmmww mw mm xw wnu m wwamwwm a . wwwmww rwá
    vwuywwmsm&wm c aiw äwwömähwbwemwnwwuamwvwwww

    \<
,       w        ?    á I,
%  .,v.&  a  å.;.*:,; 2
```

Output file 9: *Tesseract – Image* i

```
di* 5:7

      Jb  -W
l Forts från       <å::$.
Fastigheten Oxie 1:5  4.

De1 om 1 261_;% ay fastigheten Oxie 1:5 utgörwgggfällighetenz
Qångggiften Szltrz
_Areaiz 2 041 038 _mz
```

Output file 10: *Tesseract – Image* j

```
Image (shown in Appendix A)

a b C d e
Tesseract 1697% 3559% 1672% 3678% 164%
Ocrad 2896% 5793% 362% 7578% 51151%

OpenOCR 2343% 4901% 1733% 7442% 17168%
GOCR 6262% 4104% 7768% 6286% 78.53%
OCRopus 3287% 1579% 3764% 41152% 39.38%
TOCR 474% 4413% 1322% 7207% 1126%
Abbyy 9513% 29.16% 2456% 2703% 1531%

Software

Table 1: Results of the OCR scans.

so

0 60
,3


U

3 40
E
La

20

0

a b c d e
Image (shown in Appendix A)

Iii Tesseract Ii Ocrad lili OpenOCR Iii GOCR
OCROpus In TOCR M Abbyy

Figure 2: Illustration of the scan results.

The mean error value in increasing order for each of the OCR tools is shown
below:

TO CR 81169%
Abbyy 21.11%
OpenOCR 22198%
Tesseract 24149%
OCRopus 33144%
Ocrad 5007%
G 0 CR 6455%
```

Output file 11: *Tesseract – Image* k

```
Testing OCR tools

Angelica Gabasio

1 English

This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.


Figure 1: This is a cat.

2 Swedish

Och så lite svenska tecken. Båten åker i sjön.
Det ar blött.
```

Output file 12: *Tesseract – Image* l

```
OP\:i(.Q\ C\m\rm:5:e.r racoqn\ 1on

ANe1:\_\c/+\ (';{A%/\5\

2075
```

Output file 13: *Tesseract – Image* m

```
The algorithm uses a database of letters and numerals to match the char-
acters. The algorithm will calculate a matching value for each character and
selects the option with the best value.[1]

A way of improving the output is to check if the scanned word exists in
a dictionary. If it doesnt, it is likely that some of the analyzed
    letters are
wrong, and some of the letters may be changed so the word matches a word in
the dictionary. This is done by adding a penalty value to the matching value
if the analyzed character result in a word not in the dictionary. This makes
it possible for another letter to get a better matching value if the word
    with
the new letter is part of the dictionary. The letters are not always changed
    ,
sometimes a non dictionary word gets a better matching score, even with
    the
penalty value, than the one with the changed letters that is in the
    dictionary.
Even if using a dictionary may help the output accuracy it is no guarantee
that the correct word will be found.[1] An example of the use of a
    dictionary
can be seen in Figure 1.

Other ways to get a better output is to use machine learning, where the
OCR tool can be trained to recognize more languages or fonts, or to use the
knowledge of which character combinations are more likely to appear.

The better the image is, the more accurate the result gets. If the OCR
algorithm is applied on a horizontal text with very little or no noise and
    the
letters are well separated from each other it is possible to get a very good
result. But if the image is blurry, warped or noisy, or if the characters
    are
```

```
merged, it is likely that the output gets worse than it would otherwise.[1,
    2]
The output accuracy will also depend on the image resolution.

OCR software is generally better at recognizing machine writing than
handwriting. In machine writing the same character always looks the same,
given that the same font is used. With handwriting the characters differ a
    lot
depending on who is writing, and even if the same person has written a text,
the same character will be a little different each time it is written.
    Machine
writing often consist of more distinct characters, and since handwriting is
more varied than machine writing, it is harder for the algorithm to match
the characters to its database.
```

Output file 14: *Tesseract – Image* n

B.2 Ocrad

```
| ' Fartg ,__ '.-'_. .i__\ S_
rastighe_e_ 8i1en_ 4 '‘,'
. __+ . _--- _.- ____.____-.. _.___
_: Se akt ang 8ilen 4 i lagf skåp
_Del om 836 m' av rastigheten Bilen 4 utgör nybildning av
_ samräl7igheten Stapelbädden s:4 (se överenskormP7se i avyttrings- _
_ akt ang del om ca 3 ozo m' av fastigheten Bilen 4)
Areal: 339 5q6 m'
i -

_l ___,'Se akt ang Bi7en 4 i lagf skåp
__De7 om 737'mZ..aY rastigheten Bilen 4 utgör Galeasen l..
Del om Z zz9 m' av rastigheten Bilen 4 utgör Caleasen z_
__ _: 336 580 mZ "
| _
_ y G v!
| .
|

|
_ Forts från fasti héten sid _
_ Bilen 4
__k__ä_ g Bi | é_ ___-_-.îä_-___._-_--
De7 om 63 m' av samfälligheten Flaggskepparen s:l har tillagts
fastigheten Bilen 4.
_Del om 9 m' av rastigheten Bilen 4 har ti7lagts samrä)7igheten
Flaggskepparen s:l.
Areal: 33z 894 m'

_: Se aktang Bilen 4 i lagf skåp
_Del om 96 478 m' av fastigheten Bilen 4 utgör fastigheten
8ile.n IZ.
Areal : 236 416 m2
y G v!
```

Output file 15: *Ocrad – Image* a

```
_ . fartg ,,_/- ' .-__.. ___, 6 ,
_astiahe_e_ Bilen 4 ' ‘'|
____ _ --_ _-__._ _________-____ -___

_: Se aktang Bilen4 i lagr skåp
_Del om 836 m' av rastigheten Bilen 4 utgör nybi7dning av
,' samrä7ligheten Stapelbädden s:4 (se överenskomnelse i avyttrings- _
_ akt ang de) om ca 3 ozo m' av fastigheten Bi7en 4)
i Areal : 339 546 m'
|

_____;'Se akt ang Bilen 4 i lagf skåp
_ Del om 73__mZ.ay rastigheten Bilen 4 utgör Galeasen l..
Del om Z zz9 m' av rastigheten Bilen 4 utgör GaFeasen z_
__ Areal : 336 580 mZ "
i_ -
|
|
_ v c v!
| .
|

|
|
```

```
i Forts frgn fasti héten sid _
_ Bilen 4
- _: Se ak_ä.ng 8ile_____-_.-íä _r --_._p--_- __
Del om 63 m' av samfäl)igheten Flaggskepparen s:l har tillagts
fastigheten Bilen 4.
_De7 om 9 m' av fastigheten Bilen 4 har ti7lagts samräl7igheten
Flaggskepparen s:l.
Areal: 332 894 m'

_: Se aktang Bilen 4 i 7agf skåp
_Del om 9ó 478 m' av rastigheten Bilen 4 utgör rastigheten
Bile.n IZ.
Areal: 236 416 m2
v G v!

|
```

Output file 16: *Ocrad – Image* b

```
|
rOrtS _, ,---.. ij
_te_' . _ _'_'\
_ 6,

. ._ _-___h_____._...\_______.__. __._
Fr 7 re 19/8-09: Se akt ang 8i7en 4 i lagr skåp
/Oel om g36 mz
av rast7gheten 8ilen 4 utgör nybiFdning av
samrä77igheten Stape7bädden s:4 fse överenskormP_se i avyttrings- _
akt ang de7 om ca 3 ozo m' . .
_' _: J39 546 mZ av rast7gheten 877en 4)



__/_.:-  ''e ak' ang Bi_en 4 i 7ag, ,kåo

Oel om 737 mZ aY rastigheten Bilen 4 utgör Galeasen _ .
Oe7 om Z zzg mZ . .
__ _: 3J6 5g0 m_' 'a"'ghe'e" 8i'e" 4 'tg_r _a7easen z_
/


Y G v!



F_rts frgn fasti héten

__ _ä       B, s_d7
Fr | r . _.__\__ _.
De7 om 63 mZ . i e" 4 __ä __ -___p_-_____

av samfä)ligheten F7aggskePparen S:I har ti7lagts
rastigheten 8ilen 4.
/Oel om g mZ
av fast_gheten 8ilen 4 har ti7_agts samrä_ligheten
F7aggskepparen s:l.
_: 33z 894 mZ



Avst re z3/ll-_olo: Se akt ang Bi7en 4 i 7agr skap
|
```

```
Oel om 96 478 m' . . . .
av rast7gheten B7len 4 utgór rast7gheten
Bile.n IZ.

_: z36 416 mZ


v G v!
```

Output file 17: *Ocrad – Image* c

```
.|__ har\ del o_ c4 _6_ _..av__asti_h___
._ . _ ...-_ . ... _ _._ ..__ ._----__ _._-_ÖO  --- .. ..-_._-_.? .- ._?___?
    _'_ - . __ .- ____
- -B__é_4 á_ttrátš..--s_kt __ -.deí a.?_ om c_-__:  Ö6Ó:._'__í_m__
_______s_ ghe____7-_   7_i   .__.7_agf_ ?k_åp . Avst_.reg 16/3-z0_11 __
    _______ _

  ---._É.n7 _ t -.á_táí--Á 34_žóíó-   --h-a___-öm c_--ž- __Ó.-m' .á_. ras-
    t_he_é - ._?_
.__ B_|.en 4 avyttrats. __ AySt reg -15/7?10 __.___ - ---- ---.------.--
-. --- .Á _r é   --í5_íO_.-.šé._ t_-i-_.  šá_t-an_._éíóm._á --_4ÓÓm' -----
    ---
__ ._ . _ . _ _ ___ . __ ___ _ _ _ _ _ _ _ ____ _ ._ __ . _ _ av
. _ Bi len 4. _______.__ _____ __ _ -- . ----
__ _ l De7. . om z _428. m2 .av_ __e_ -_u_t ._i_________-Klyvaré  - j - Á
    .34lž-o-í
_-.lDel-.om l 31Z mZ _ av 8i-|_e_ 4. u?gör__ras.tigheten .Kly áré - ž
    %-.---- --..----
| -. --- -
.__ Areal : . 33? 8q0 mZ _____ __ _.. __ ________ _________ _____. ______
    ___ ____
?. _ ____ _ . ._?_____________ _____ ?. _ ____ _.__ ._ _i._______.
    _ !  _________ ._. _  For_s. s_ d   7
```

```
_: Se aktang 8ilen 4 i 1agf skåp
l Del om l 017 m' av Bilen 4 utgör Klippern l.
|Del om 813 m' av Bi7en 4 utgör Klippern z.
l Del om 631 m' av Bilen 4 utgör K7ippern 3.
l Oe_ .om l 078 m' av Bilen 4 utgör Klippern 4.
f Del om z z55 m' av 8ilen 4 utgör _osterbåten l.
l Del om l 627 m'. . ..    .
t Del om l 7g4 m2 __ B!'e" 4 "g  ' Wsterbaten z
Bllen 4 utgör Koggen l.
| Del om l 666 m' av Bilen 4 utgör Koggen z.
Areal: 2z5 545 m2

_: Se akt ang delar om ca zo6o m2 resp 11 m2 av
fastighéten Bilen 7 i lagf skåp
_D_|-om 759 mZ av Bilen 4 utgör Bilen 13. A 515 och F 11
. Areal: zz4 786 mZ z o o 9
```

Output file 18: *Ocrad – Image* d

```
_agtigbeten Oxie 1:5
_r l re 1 4-_9: se akt aAg Pastigh Oxte 1:1 m fl
Det om 10 801,7 _ a_ 9amPälligb_ten Oaie s:2 har tilLagtg
Pagttgbeten  _ie  ':_ 'g _q/'íg_g'
```

```
_ rt __ 5p,__ __ ___..
__ , .

Enl avtal _ har Zel av fastieheten Oxie 1:5 avyttratsl
l9_9
SE ak_ an_. c'd T8.ooo l_ av fa_tigheten Oxie 21;2. Delen har
vid rrgl. reg. 15/1-81 visat sig innehålla T.2g7j9 _.
_: 2 227 292,6 _
En . _ ,'trd_ il 2 / 9_1 Ehr ._l on, __.bl_, y ___ TV OyLi_ 1:5,
    avs.
t
T._._ _i I__ fas__S_c-ten E_rnl äf,lni_berl l _nvT/L.i_a_s. .'_'__ "_a'
. _ . _ _ _
Av9t reg l_/3-88: se akt ang Oxie 1:1 m fl
I.el om l_ 225 m_ ay Oxie l:5 uteör rornl_mningel_ 'L
_: 2 l8o _s2 m
```

Output file 19: *Ocrad – Image* e

```
. . bn __Sy
l .
Fz,stighete_ Oxie 1:5
Biláä_ _eAom __anläggniN6 _v f_stigh_tErna Oy-ie 1:1,
GxiP, IT:2, Oxi_ _9:íD OyL_e 20:8, Oxie 25:9, Oxie _8:v och
O-rtiE 29.:1.
_ 2.231.22__

.Särskilt._na_n_ enl.. reg. besl?t. den .1S/_?92: Fr.e.d.?_.ksg_e%hg
Jqr _. re .17 íO-r : se akt. rvg . stg 2440 m.fl. i rosie
Dpl_r cIn 85,G __ _.v s_nf_l_d_. _mr__E_ .n__, 1_956_0 _ _ . __ -
_ _ . __ c_m
f__bh_ omr___i _hr__ s:b. __ _o?,_ T% __v _.t%Zsäg_n. 2E__? _ _ l
, .
__.___,4 ___ _v s_____äL _ _ ___ST,2_6g _z,r. fiV__pcts r__t_C_%t_n
_I___ _: _.
```

Output file 20: *Ocrad – Image* f

```
ror__ f_ån
Fzs_ig_eten _ _ e __5 4.
__| om _ Z_____?__ast___eten_ Ox____l_5 utgör___amräll__g_eten_ ___ _ _ _
_-_____-__--__ __ _ - ---- _ -- ----------_- _ -----. _-- ___ --__
    ---  - _ _
___t á_7__A___/_Óo6_-_____- Óm___ ž_____- m' a-v ._aštig_-té -- - - -_
.____e__ _5 avyttrats _ ____ ____ .____ __ _ _i___ - __- _ - _ _ !_-- _ - _
    .-_- __ __ -- _ - -__ .
```

Output file 21: *Ocrad – Image* g

```
ror__ f_ån
Fzs_ig_eten _ _ e __5 4.
__| om _ Z_____?__ast___eten_ Ox____l_5 utgör___amräll__g_eten_ ___ _ _ _
_-_____-__--__ __ _ - ---- _ -- ----------_- _ -----. _-- ___ --__
    ---  - _ _
___t á_7__A___/_Óo6_-_____- Óm___ ž_____- m' a-v ._aštig_-té -- - - -_
.____e__ _5 avyttrats _ ____ ____ .____ __ _ _i___ - __- _ - _ _ !_-- _ - _
    .-_- __ __ -- _ - -__ .
```

Output file 22: *Ocrad – Image* h

```
__t__ ____ _
_____ghPt&_ k____ __4 _l,
_‘ ____,_ ____m _____á_’ _________^_’ ____’ _____’ _______’‘’_
e&n,R___’, n.gQw _: a . _ .__. , , _
_ _ .- ___ _ ___ __ _ _ _ . __ -_ _ . _
_____ _ E _G_, _ma_ _a _ . _ _ _ . __ _ . . .

__‘g___g_,___k ___ma_ RQr _________& 2q ,_ _’ &_ p__š__hRg_R
OW_g _ _S &v__er__2_. , , _ , _ _ , _, _. ,, ___ . . . _ __ ’

. _ _ _ .’ _ _ _ . _ ’ . _ . . ’ _ _ - - _ _
. _ _ _ _ _ .
```

Output file 23: *Ocrad – Image* i

```
. o’>___._
ror__ fr å_ oo ___ ’
Fzstig__t__ __íe _:5 __‘_.
\
4.

_ De_om l Z6_______v_astighe_n O_ie 1:5 utg_r___am_áll__g_ete_____ __ _

_G__qqr_r__ S: l ._____ ___ _____ _ _ __ _____ _____

_Areal: 2 _l 038m’ _____ ._____ __ .__ _ ___ _____ _ .
.
--_-7 igt-.-á ta7_ 32 l_ óó6 -_____ _--3 0 _-.á r-- 7gheté_

_ __ _5 _vyttr_s_____.. ___ _____ ___ _____ - _ _ -_ _.----- ___ -_._
- - ?__ ..

:‘- -- --- --_-.- .-- -----. _ - --. -- ----._-- ---_- t -_---
- __ - - __- _ _

_ _ _ ‘_-_ _ - _ ----- _ __- . . - - _____________ -. - _____-
_____-_ - _ | _ ----_- _ _ _ _ -- _

-.- --. _ _ -_-.-- .-- _ ---_-- - -- _ -..-- - --_ - .---__- _ ----- -
--.-
```

Output file 24: *Ocrad – Image* j

```
Im_ge l,_hown _n Appen__ A)
_ h L d _
_ T_____r_Lt 16.97% y5.59% 16.72% y6.7_% 16.4%
_ OLr_d 2_.96% 57.9y% y6.2% 75.7_% 51.51%
_ Op_nOCR 2y.4y% 49.01% 17.yy% 7.442% 17.6_%
_‘ _OCR 62.62% 41.04% 77.6_% 62._6% 7_.5y%
OCRupu_ y2._7% 15.79% y7.64% 41.52% y9.y_%
TOCR 4.74% 4.41y% ly.22% 7.207% 11.26%
AhhW 9.51y% 29.16% 24.56% 27.0y% 15.y1%


T_hl_ l: Re,_TILt,_ of the OCR ,___n,_.
```

```
_u



_ _u
w

h
A 4u
h


2u



u
_ h L d _
Im_ge l,_hown _n Appen__ A)


_& T____r_Lt ll OLr_d lll Op_nOCR _| _OCR
Oo OCRupu_ || TOCR ll AhhW

Figur_ 2: ILLTl,_tr_t_on of the ,___n re,_TILt,_.

Th_ m__n _rrur v_lu_ in inLr___ing urd_r fur __Lh uf th_ OCR tuul_ i_ _huwn
h_luw:

TOCR _.169%
Abbyy 21.11%
OpenOCR 22.9_%
Tesseract 24.49%
OCRopus yy.44%
Ocrad 50.07%
GOCR 64.55%



6
```

Output file 25: *Ocrad – Image* k

```
Testing OCR tools

Angelica Gabasio


l English

Thi_ i_ a phge with a picture to test the OCh tools.
Thc pagc conthins tcxt in Engli_h hn_ Swcdish.
```

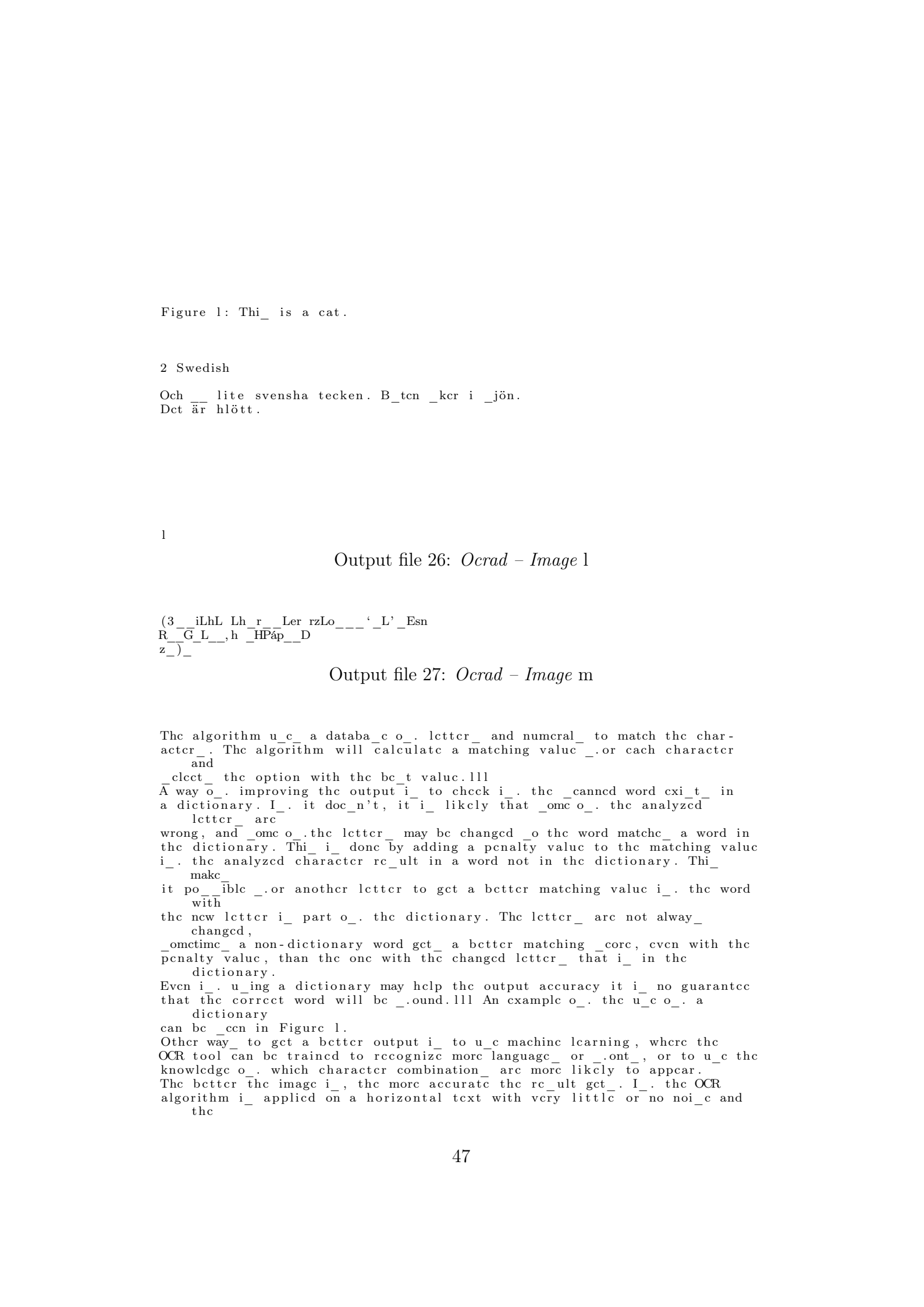


Figure l: Thi_ is a cat.

2 Swedish

Och __ lite svensha tecken. B_tcn _kcr i _jön.
Dct är hlött.

l

Output file 26: *Ocrad – Image* l

(3__iLhL Lh_r__Ler rzLo___‘_L’_Esn
R__G_L__,h _HPáp__D
z_)_

Output file 27: *Ocrad – Image* m

Thc algorithm u_c_ a databa_c o_. lcttcr_ and numcral_ to match thc char-
actcr_. Thc algorithm will calculatc a matching valuc _.or cach charactcr
and
clcct thc option with thc bc_t valuc.lll
A way o_. improving thc output i_ to chcck i_. thc _canncd word cxi_t_ in
a dictionary. I_. it doc_n’t, it i_ likcly that _omc o_. thc analyzcd
lcttcr_ arc
wrong, and _omc o_.thc lcttcr_ may bc changcd _o thc word matchc_ a word in
thc dictionary. Thi_ i_ donc by adding a pcnalty valuc to thc matching valuc
i_. thc analyzcd charactcr rc_ult in a word not in thc dictionary. Thi_
makc_
it po__iblc _.or anothcr lcttcr to gct a bcttcr matching valuc i_. thc word
with
thc ncw lcttcr i_ part o_. thc dictionary. Thc lcttcr_ arc not alway_
changcd,
omctimc a non-dictionary word gct_ a bcttcr matching _corc, cvcn with thc
pcnalty valuc, than thc onc with thc changcd lcttcr_ that i_ in thc
dictionary.
Evcn i_. u_ing a dictionary may hclp thc output accuracy it i_ no guarantcc
that thc corrcct word will bc _.ound.lll An cxamplc o_. thc u_c o_. a
dictionary
can bc _ccn in Figurc l.
Othcr way_ to gct a bcttcr output i_ to u_c machinc lcarning, whcrc thc
OCR tool can bc traincd to rccognizc morc languagc_ or _.ont_, or to u_c thc
knowlcdgc o_. which charactcr combination_ arc morc likcly to appcar.
Thc bcttcr thc imagc i_, thc morc accuratc thc rc_ult gct_. I_. thc OCR
algorithm i_ applicd on a horizontal tcxt with vcry littlc or no noi_c and
thc

```
lcttcr_ arc wcll _cparatcd _.rom cach othcr it i_ po__iblc to gct a vcry
    good
rc_ult. But i_. thc imagc i_ blurry, warpcd or noi_y, or i_. thc charactcr_
    arc
mcrgcd, it i_ likcly that thc output gct_ wor_c than it would othcrwi_c.ll,
    21
Thc output accuracy will al_o dcpcnd on thc imagc rc_olution.
OCR _oWwarc i_ gcncrally bcttcr at rccognizing machinc writing than
handwriting. In machinc writing thc _amc charactcr alway_ look_ thc _amc,
givcn that thc _amc _.ont i_ u_cd. With handwriting thc charactcr_ diKcr a
    lot
dcpcnding on who i_ writing, and cvcn i_. thc _amc pcr_on ha_ writtcn a tcxt
    ,
thc _amc charactcr will bc a littlc diKcrcnt cach timc it i_ writtcn.
    Machinc
writing oWcn con_i_t o_. morc di_tinct charactcr_, and _incc handwriting i_
morc varicd than machinc writing, it i_ hardcr _.or thc algorithm to match
thc charactcr_ to it_ databa_c.
```

3

Output file 28: *Ocrad – Image* n

B.3 CuneiForm

```
Forts
Fastigheten, Bilen 4 '
Fr 1 re 19/8-09: Se akt ang Bilen 4 i lagf skåp Del om 836 m av fastigheten
    Bilen 4 utgör nybildning av
2
samfälligheten Stapelbädden S:4 (se överenskommelse i avyttringsakt ang del
    om ca 3 020 m av fastigheten Bilen 4)
2
Areal : 339 546 m
2
iå
Del om 737 m2 av fastigheten Bilen 4 utgor Galeasen 1.
Del om 2 229 m av fastigheten Bilen 4 utgör Galeasen 2
2
Areal: 336 580 m2
Fo~ts från fastigheten
Bilen 4
Fr 1 re 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m av samfälligheten Flaggskepparen S: 1 har tillagts
2
fastigheten Bilen 4.
Del om 9 m av fastigheten Bilen 4 har tillagts samfälligheten
2
Flaggskepparen s: 1.
Areal: 332 894 m
2
Avst re 23/11-2010: Se akt ang Bilen 4 i lagf skåp
av fastigheten Bilen 4 utgör fastigheten
2
Bilen 12.
Areal: 236 416 m2
```

Output file 29: *CuneiForm – Image* a

```
For ts
Fastigheten Bilen 4 ' '
1
Del om 836 m av fastigheten Bilen 4 utgor nybildning av
2
samfälligheten Stapelbadden S:4 (se överenskommelse i avyttringsakt ang del
    om ca 3 020 m av fastigheten Bilen 4)
2
Areal: 339 546 m
2
I
Del om 737 mZ av fastigheten Bilen 4 utgör Galeasen l.
Del om 2 229 m av fastigheten Bilen 4 utgör Galeasen 2
2
Areal: 336 580 mZ
Forts från fastigheten
Bilen 4
1
Del om 63 m av samfälligheten Flaggskepparen S: 1 har tillagts
2
fastigheten Bilen 4.
Del om 9 m av fastigheten Bilen 4 har tillagts samfälligheten
2
Flaggskepparen s: 1.
Areal: 332 894 m
2
Avst re 23/11-2010: Se akt ang Bilen 4 i lagf skåp
```

```
Del om 96 478 m av fastigheten Bilen 4 utgör fastigheten
2
Bilen 12.
Areal : 236 416 mZ
VGV!
```

Output file 30: *CuneiForm – Image* b

```
Forts Fastigheten Bile ~ 4 Fr 1 re 19/8-09: Se akt ang Bilen 4 i lagf skåp
    Del om 836 m av fastigheten Bilen 4 utgör nybildning av samfälligheten
    Stapelbadden S:4 (se överenskommelse i avyttringsakt ang del om ca 3 020
     m av fastigheten Bilen 4)
2 Areal: 339 546 m
2
/- Del om 2 229 m av fastigheten Bilen 4 utgör Galeasen 2
2 Areal: 336 580 m2
VGV! For ts frå n fastigheten Bilen 4 Fr 1 re 30/7-10: Se akt ang Bilen 4 i
    lagf skåp Del om 63 m av samfalligheten Flaggskepparen S:1 har tillagts
2 fastigheten Bilen 4. Del om 9 m av fastigheten Bilen 4 har tillagts samfä
    lligheten
2 Flaggskepparen s:1. Areal: 332 894 m
2 Avst re 23/11-2010: Se akt ang Bilen 4 i lagf skåp Del om 96 478 m av
    fastigheten Bilen 4 utgör fastigheten
2 Bi 1 en 12. Areal : 236 416 m2
VGV!
```

Output file 31: *CuneiForm – Image* c

```
Enli t avtal A 515 och F 11 har, del om cp 760 m av fastigheten
2
2009
akt ang delar om ca 2 060 m resp 11 m
2
lagf skåp. Avst reg 16/3-2011
Bilen 4 avyttrats. Se
av fastigheten Bilen 7
Enligt avtal A 34/2010 har del om ca 2 400 m av fastigheten
2
Bilen 4 avyttrats. Avst reg .15/7-10
Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m av
Bilen 4.
( Del nm 2 a2B m2 av Bilen a utgör fastigheten Klyvaren l. a 3a/2Dl
Del om 1 3]2 m2 av Bilen 4 utgör fastigheten Klyvaren 2.
Areal: 332 840 m2
Forts sid...7
Se akt ang Bilen 4 i lagf skåp
4 utgör Koggen 2.
Areal:
Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skåp
Del om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 1]
Areal: 224 786 m2
200 9
I Del om j Del om j Del om ~ Del.om I Del om ~ D.l om t Del om l Del om
1 017 m av Bilen
```

```
2
813 m av Bilen 4
2
631 m av Bilen 4
1 078 m av Bilen
2 255 m av Bilen
2
1 627 m av Bilen
1 784 m av Bilen
1 666 m av Bilen
2
225 545 m2
4 utgör Klippern l.
utgör Klippern 2.
utgör Klippern 3.
4 utgör Klippern 4.
4 utgor Kosterbåten 1
4 utgör Osterbåten 2.
4 utgör Koggen 1.
```

Output file 32: *CuneiForm – Image* d

```
Fastigheten Oxie 1:5
Pr 1 re 1 4-79: se akt ang fastigh Oxie 1: 1 m fl
Del om 10 801,7 8 av samfälligheten Oxie s: 2 har tillagts
fastigheten Oxie 'I: 5 (W 29/197))
~'. X 39 x8'~

A61 260
Enl. avtal ]979 har åel av fastigheten Oxie 1:5 avyttrats,
se akt ang. ca 78.000 uf' av fastigheten Oxie 21;2. Delen har
vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 m .
areal: 2 227 292.6 9)
Enl. avtal A 20/1981 har del om ~ 2,615,9 rn' av Oxie 1:5, avs.
att bilda f as+9
t-gheten Pornl ämningen 1 avvttras. ( ~<å', . -4y')
r eg 17/3-88 : se akt ang Oxie ] : 1 m
Bel om 14 225 m ay Oxie 1:5 utgör Fornlämningen 2
2 180 452 m
```

Output file 33: *CuneiForm – Image* e

```
Fastigheten Gxie 1: 5
Bildad genom s~~anläggning sv fastigheterna Qxie 1:1,
Gxie 17: 2, Oxie 19: 1, Gxie 20: 8, Oxie 25: 9, Oxie 28: 1 och
Oxie 29:l.
'Areal efter sa~anlä- rin : 2.251.222 rS
Särskilt namn enl. reg. beslut den 'l5/4-92: Fzedriksbe~g
Fr 1. re ~ 17 10- : se akt ang, stg 2440 m fl. i Posie Delar om 85,0 rE av
    samfällda område sdj, 1.956,0 m' av samf ällds området Qxie s : l.
    "~.507, 2 iP, av stadsägan, 28 4 och 22.122 4. ;E av s+adsägar 2467,
    2468 hsr tillagts f astigheten
Qx" 1: 5.
```

Output file 34: *CuneiForm – Image* f

```
For'.s från
Fastigheten Oxie 1:5
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
```

```
Gånggriften S:1.
Areal: 2 041 038 m


Enligt avtal A 329/2006 har del om ca 23 300 m av fastigheten
2
Oxie 1:5 avyttrats.
```

Output file 35: *CuneiForm – Image* g

```
For'.s från
Fastigheten Oxie 1:5
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m


Enligt avtal A 329/2006 har del om ca 23 300 m av fastigheten
2
Oxie 1:5 avyttrats.
```

Output file 36: *CuneiForm – Image* h

```
. E@t'$94 4~$ A 329/2~ 540 4@3 ~ C4 23 ~ S 4V f@S4$9W'W5
QX)e I:5 eyyttee45 ,.
```

Output file 37: *CuneiForm – Image* i

```
Forts från
Fastigheten 0xie i:5
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
Ganggriften S:1.
Areal: 2 041 038 m

Enligt avtal A 329/2006 har del om ca 23 300 m av fastigheten
2
Oxie 1:5 avyttrats.
```

Output file 38: *CuneiForm – Image* j

```
l,)1)li' 1; Ii'r i(1t i/ t/ir (/CIi'  rn
llti nti,nt it(in r ilni ut nti ti,i nt:, inili t lin i,   lt i)l tlti
    (/('1( ti)i)l t l nt
lii lint:
'J'OCB ebb> r OpeuOCB 'J'enner tct OCBopian Ocr tcl COCB
```

```
1 t,',nti' ': I11  tint!i(n ii/ t/ir  rn  rr i(1t .
   . 1 ()')' ')1.11' ') l. 1')' ,),). 1 1' )(l.(l f) l. l l
```

Output file 39: *CuneiForm – Image* k

```
Ti uti o ()C.'H tOOls
Angelica Gabasio 1 English Tlus (s,( p " (' u(l h,( ln( I ur(' I() I( st I
    h(' OOB louis. Th( p,( ( (uul,(n(s I(sl u( Eu hsh,uul Su((hsh.
Figure D This is a cat. 2 Swedish Och så lite svenska tecken. Båten åker i
    sjön.
Det är blött.
```

Output file 40: *CuneiForm – Image* l

```
p()<g( 4h.-pro,eke~  c >anile-n
VJ

Ail
GE.Ui ~A, C>HEW& i 0
```

Output file 41: *CuneiForm – Image* m

```
) ili',Ii"r)llti1111 IL i', ,I rf,lt,lf),1, i' r)i ii'tti'I, Illrf 1111111i'
    I IL tr) 111 Itr il tili' I il 11â       ,ii tr I, . ')'llr ,ii," r)1
    itliiii il ill i,ili iil,itr ,i Iil,iti lliii," v,iiiir ir)1 r,ii ll i
    il,il,ii tr I ,iiiif , I'lr'i t, till' r)f)tlr)ll Ll itll till I)r, t v,
    tltti .[i ]
. ) Ll IL r)i 1111I)li)L 111,'~ till' r)litf)lit I, tr) I ilr'I il li till',
     I 111111'rf Lli)lrf 1'Xi, t, 111 I rflr t l))11 11 LI ii It rfr)ii 11 t
    . It 1 il)Il'il til It , r)llli' r)i tili' 111 til /I'if ii't ti I, Ili'
     Ll Ii)il~.,tilif, r)1111' r)i till' ir't tr'I, 111,IL I)r' I i1,111" I'
    rf, r) till' Lli)lrf 111,1tr ilri ,I Lli)lrf 111 till rlir t i))ii,ti )I
     'l lli, i, rlr)111 I)y ,trlrliii," ,I f)r Ii,tlt y v,tltti tr) till Iil
    ,ttr lliii," v,tltti Ii iili' 111 til /I'rf I il 11 ir ii'I I il Iiii
    111 I Lli)I rf 11))i 111 iili' rflr il))11 11 \. ) 111 111 lilii it f)r)
    ,, il)lr ir)1 ,tiir)till I ir ttr I tr),"I t ,i I)r t tr I 111 itr lliii
    ," v iliir ii till Ilr)1 rl Ilitil tili' Ili Ll ii t ti I 1 I) 11 t r)i
    tili' rflr t l))11 IIL. ) ili' ii t ti I, Ili' 11))t tiil IL, I il 111~1
     if. , r)1111'tllllr', ,I 11))11-rflr t l))11,11 I Lli)I rf "I't, ,I I)r
    't tr'I 111,1tr i1111" , I r)lr'. ltL1'11 Llltil till' I)i'11 lit l \
    lilii'. fil 111 tili' r)lli' Llltil tili' I il tll ~i'if ii't ti'I, fil
    It 1 111 tili' rflr tlr)11 II I . L L1'II li il, III," I rlir tir)ll,il
    \ Ill i) lli'if) tlli' r)lit f)lit ii i ill ii \ it i, llr) " il,il
    illtr'I' tll,it till r r)1 11 r t Ilr)1 il 11 ill I)r ir)ttilrl.[l] . 41
     I x,tiilf)li r)i till ii, I r)i ,i rlir t i))ii,ti y
ill I)r', I'I'II lll i'i illr' l.
( )till I Ll IL, tr) "I t I I)r t tr I r)lit f)lit I, tr) il I' Ill ir
    llllr' lr' tl llllll ~. Krill lr' till' ( )( l) tr)))l i,ill I)r' tl
    illlr'il tr) lr'i)),"  lll/I' Illr)lr' i ill," il I,"ri r)I ir)lit, . ))I
     tr) il, I' till' lill))till'il," I' r)i Llllii ll I il il ir tr I I r)
    lill)lll itl))ll, il r' Illr)lr' lilll'i) tr) if)f)r' Il.
) ili' I)i't ti'I tili' 1111,1" i' I, . tili' 111))li' ,Ir I Iil,lti' tili'
    Ii', Iiit " i't, . ii tili' ( )( i) tl,'~i)lltllllll I, if)f)llr'rl r)ll
    I llr)l I/r)lit Il tl'x t Ll ltll L1'I I lit tlr' r)I llr) llr)l, I'
    illrl till' lr t tr I ilr' Kil il I f) il itr rl il))lll I' ir ll r)till
     I lt I f)r),, ll)lr' ti) ,'~l t I L1 I I ,'~i)i)il lr', illt. f fttt li
     till' lill I" I' I, I)lttl I I . Ll il f)r'rl r)I llr)l, LI r)I li till
```

```
    ' r ll il ir tr'I, il r' 1111 I "I il. it i, lilll ly t il it till r)
    iitf)iit "I t, ilr)1 ~1 tll tii it ilr)ttlil r)till I il i, I .~l. 2~ l
    lll ' r)lit f)lit ,ii i Ill,ii \ Krill ,tl, r) ilr'f)r'llil r)ll till'
    lill I"I' lri r)illtl))ll.
( )( l) r)ittl il r' I ,"I llr I Ill) I)r'ttr I it lr r r)," lll/ lll," Ill
    Ir llllr ' Lll itiii," t il tii ll illrftl lltlll ~. lll Ill ir llllr '
    LII ltlll ~ till', illlr ' r ll il ir tr'I tltl IL, lr)r)ll, till', illlr
    '. ,'~111'11 til It tili', 1111i' ir)lit 1 Ii, i'rf. A Itil il lllrftli
    ltill ~ tili ' I il 11 Ir ti I, rfliii I I ir)t rlr f)l llilll~ r)ll
    Lill)) I, Ll lltlll~. Illrl ltL111 li till', illlr ' f)r I, r)ll ll I,
    Llllt tr II I tl xt. tili',,1111i' I il,ll,lr ti I Lllii I)i',I ilt tii'
     rfliii Ii'Ilt i',Ir il tllli ' It 1 LII ltti'11. L i,tr 111111' LII
    ltlil ~ r)iii'11 i i)11, I, i r)i 111))li ' ifl, illli i i il 11 ii ii'I,
     . Illif , Illi i' il tiliftl Iltlil" I, Iilr )I I II ti il il t il, tii
     111, I 1 lliiil il I i till," . i t i, ll, ti 1 ll I ir )I till , tl,"
    r )I i tillit t r ) 111, I t 1 ll till ' r ll il ir tr'I, tr) lt, rl it
    il),i, I'.
```

Output file 42: *CuneiForm – Image* n

B.4 GOCR

```
__D__DD_eeeol_r_tooom_mm296_3r2_m_a2_m9m      _    99    _     99     t
    rGa_e_asen/__2____a9ts
_ _ _x \                                  _
fastigheter1 _i1en 4

_Fr 1    19/8 09:  Se akt ang Bi1en 4 i  1agf skåp
2 av fast_.gheten Bl.1en 4 utgo..r nyb_.1dn_.
samfälligheten Stape1bädden S:4 (se överensko nme lse i avyttrings-
2 av fast_. heten Bl.
2


A _t r   30/9 09:  Se akt ang Bi1en 4 i  1agf skåp
Del om J3J m2 av fastjgheten Bjlen Q utgör Galeasen l
2 av fast_. _eten B_.len 4 utgo..
Area1;  336 580 m2
_                                              V G V!



8, fastl_ _ete_                              Sid J
Bi1en 4
_fr 1  r   30/7 10: Se akt ang Bi1en 4 i  1agf skåp
2 av samf.a.111. heten F_a  skepparen s.1 har t_,
fastigheten Bi1en 4.
2a
V faStl9hete_ 8l1en 4 har tl11a9tS Samfäll19heten
flaggskepparen s:l.
2


_A st r   23/11 2010:  Se akt ang Bi1en Q i  1agf skåp
De_ om g6 4Jg m2 av fast_.ghete, B_tle, 4 ut o.., f,st_t
Bjlen 12.
Area_; 236 Q16 m2
V G V!
```

Output file 43: *GOCR – Image* a

```
__)___    t____D_ADDerele_ao_ommt_ 236332
    __2_2__m989m4avmasvamffta_as__t_____99hheettee_nnBF___lae9n9s
    Qkeuptp_9ao__r FenGas_,__elashean_r__r2t_______a9ts
f _ r t        ,, _      _' __                          ,
f__tighet__ _i_en 4

_fr 1    19/8 09:  Se akt ang Bi1en 4 i  1agf skåp
2
samfä11igheten Stape1bädden S:Q (se överensko nme 1se i avyttrings-
2
2


A _vst re  30/9 09:  Se akt ang Bi1en 4 i  1agf skåp
Del om J3J m2 ay fastjgheten  Bilen 4 utgöF Galeasen l
2
Area1:  336 580 m2                          /

_                                             v G V!
```

```
'      __rts  _rán _ast_g-hetgn                          S_d J
Bilen 4
_fr1    30/7 10: Se aktang Bilen 4i  1agfskåp
2
fastigheten 8i1en 4,
2
el Om 9 m  aV faStl9heten B1le, 4 haF tl11a9tS Samfalll9heten
flaggskepparen s:1.
2


_Avst re  23/11 2010;  Se akt ang Bilen Q i 1agf skåp
0el om g6 4Jg m2 av fast__g_eten 8__le, _ ut o.., f,,t__
8ile.n l2
Area1; 236 416 m2
V G V.!
```

Output file 44: *GOCR – Image* b

```
_________      9                t9o__9F9Ga7peapsenl_r9_a9ts
__rt_
fas_í g h e t e J1 _ i _ e _ _ 4

Fr01  re_ 1 9 / 8- O 9:  S e a k t a n g ß i 7 e n 4 i  1 a g f s k å p
0e7 ,m 8 3 6 m2 a v f a s t _ g h e t e n g i _ e n 4 _ t g ö r n y b_ J d n
      n  a v
samfä7li g h e t e n S t a p e 1 b _ d d e n S: 9 ( s e ö v e r e n s _ o
    nme l s e i a v y t t r i n g s
a_t an g  d e _ o m c a 3 o 2 o m 2 a v f a s t i g h e t e n g i 7 e n q)
A,e a 1 ..  3 3 g 5 4 6 m 2


Avst re_ 3 O / 9- O 9:  S e a k t a n g 8 i 1 e n 4 j  l a g f s k å;
De_ om 7 3 7-' m 2 a y f a s t_. g h e t e n B _._ e n q u
De7 om 2 2 2 g m2 a v f a s t i g h e t e n B i _ e n g u t ö _, G a 7 e a s
     e n 2
Area7:  336 5 8 O m 2                          /  V G V !



Fort_  _rån f a s t _ g h e t e n                        S i d J
8ilen 4
fr07  re0 30 / 7- l O: S e a k t a n g 8 j 7 e n 4 j  l a g f s _ å
Del om 63 m2 a v s a m f a__ _ 7 7 g h e t, n F 7 a g g s k e p p a, e n s..
     l h a, t_ _ 7 7
fastigheten 8 i 7 e n 4.
De_ om g m2 a, f a s t i h e t e n B 7 _ e n q h a, t 7 l _ a t s, a m f a _
     _ 7 7_ . h e t e n
Flaggs_ep par e n s: l ,
A,ea7..  3 3 2 g g 4 m2

Avst re0 23/ll - 2 O 1 O:  S e a _ t a n g B j l e n 4 j  1 a g f s k a '
De7 om g6 47g m2 a, f a s t_ _ g h e t e n g_ _ _ e n q_
 8ile.n l2.
 Area7; 236 4 1 6 m2                              V G V. !
```

Output file 45: *GOCR – Image* c

```
t_____DDDDDeeeee_____ ooooommmmm286ll3loo2315_J1mm857mmm2     ____  _  _
     _____pp__ppepepprpFeennrr32n_n_____4tl__e__2nol_o9_____  _____  _____


; _En_i0t avtal A 515 oe_ F ll  hac_ de_ Q_ c_ J__ m2 av _ast7_
2009
Bjlen 4 avyttrats.  se akt ang de_ar om ca 2 o6o m2 resp _1 m2
aV faSti9heten Bi1en 7 i  1agf skåp. Avst reg 16/3-2ol_

En_jgt avta_ A 34/2olo  har de_ om ca 2 4oo m2 av fast_.
8i1en 4 avyttrats._Avst reg 15/J-1o  ___

Avst reg 15/J-lo;  se avyttrjngsakt ang de_ om c, 2 4oo m2
Bilen 4.
De1 om 2 Q28 m2 av Bi1en Q utgör fastigheten K1yvaren l. A 34/201
De_ om l 312 m2  av Bi1en 4 utgör fastigheten K1yvaren 2.
Area1: 332 8QO m2
_  --_  - _  -      _ _    - _  _  __      _   _ -   forts sid J



_n  t    2_/1 __1_:  Se akt ang Bi_en 4 i  1agr _ká_
2 av B_._en Q utgo..F K1_.
2 av B_._en _ utgo..r K__.
2 av 8_.1en _ utgo..r K1_.
2 av B_.len Q utg.o.r K1_.
2 av B_.len 4 utgo..r  _osterbao
2
Del om 1 627 m  av Bi1en    utgör  _sterbåten 2,
Del om l 78Q m  av 8i1en 4 utgör Koggen l.
De_ om l 666 m2 av 8___en 4 utgo_., Koggen 2
Are,_,. 225 545 m2
Avst reg l6/3-2011:  Se akt ang de1ar om ca 2060 m2 resp ll m2 av
fastigheten Bi1en 7 i  1agf skåp
D_'1 om 759 m2 av Bi1en _ utgör Bilen l3. A 5l5 och f ll
Area1: 224 786 m2
```

Output file 46: *GOCR – Image* d

```
_______n_____o _____2_Tt__r22___7_g__l_62l29o_/t___665o_e2_l_ t_____  o_t_e_
     o_l __rtJ_ t6l_r_ _c Jt _l_r_l  3v o_____ rv_ _____t  avst)
_&stigbeten O_ie _ ; 5
0 _ _e_0  1 0 4_7g;  ge abt 8Dg _8stigh Ox te 1 : 1 _ _ _

_ae_1_beten o__' e  t: 5  (A_ 22_/l,_ ___)
-__ __ ____ !f ___ i,ì  ___  /_;,

E___ aY___  _gjg  har _el aV fa5_i_he_e_ OX_e l_5 aYy_tr__t5,
s_ _k__ a___,.  ca 78_OO_ _ _ av fa5tigheten Ox i_  2 lj 2. De _e_ ha_
vid f_gl.  _eg.  15/l-81 visat sig in_eh å_la 7. 2 _ 7_ 9 _


a_,-___v ___ ___'5c_ f_s-__w_h__ten _,,,_ _m,,_  ,,,.e, _, w__v__,_._c__ats_
     _dy _/'__/,  _ /  _ _ _ /
AV9t  reg  l7/3-88:  se  akt  ang oxie  l;l m f l
__.  2  l8o  452 m
```

Output file 47: *GOCR – Image* e

```
______p_y__e_____ar_ ___m___g_v5r__s_t___n_2___&
    ___l__3__nrtt_____f_t__l__lyln_lD___tr__
    _2____2__tltol_____2_t_2_______2e__ t  l_9____o _r_17_ __6___?___m0_y

_'astigh_e__e_u _xje  __ _
gi_a_a__ _e___ _-,_2n_ägg_i__ _wv f3s t ig he te_n2  0_ je __ _,
_xie_  _?:2t  _Xi_ _9, _î t  __i_e  2__8t  0X_' e  25_9t  OXie  _8__ _C__
_-_j_  29:_.

Särskilt na_mn enl.  reg.  beslut d_en 1_5/4-92:  f_____ed__r_ __.__sb._e_r._
_,__o__  _e_0___ io-_'    :  se akt,  ___g_  s_g ___4_ __?___  i_ F____D_

f_j___  om_å__?__.  ___=__ ___.,  Z,__ _4,_7?_/ r_.  _;__ _t__säe_n  2____'?
      J_,i_
___,_ _____?x ,4 ,-__?,  c_v _t_,n_,_____'_____ J___6?.__,__6_ _5_J
    ti_____n__t___  _3,5_ _icn___,t___
?5__i_  _;5,
```

Output file 48: *GOCR – Image* f

```
___AE_rea_____2_o____4__l___o_38___m_ __  _____  _____  ____  ___
 or__  f_án
f2stiy_ ,___n _ _e ì:5                                    4.
_e1  om l 261 m  av ^_astighet_en_Oxie l:_5 u_t_gör sam_fäl1 gheten
Gång_griften S _:l.
2 /

                                                                  2
n1l9t aVta1  A 329 2006 haf de1 Om_Ca 23 300 m  aV faStl9_heten
oxie l:_5_ _avytt_ats.        _          '_
```

Output file 49: *GOCR – Image* g

```
___AE_rea_____2_o____4__l___o_38___m_ __  _____  _____  ____  ___
 or__  f_án
f2stiy_ ,___n _ _e ì:5                                    4.
_e1  om l 261 m  av ^_astighet_en_Oxie l:_5 u_t_gör sam_fäl1 gheten
Gång_griften S _:l.
2 /

                                                                  2
n1l9t aVta1  A 329 2006 haf de1 Om_Ca 23 300 m  aV faStl9_heten
oxie l:_5_ _avytt_ats.        _          '_
```

Output file 50: *GOCR – Image* h

```
_____tv____  r_____r_\t_____> _n_____t_____r_____nt_____4n>
    __t_____m  l_
    ___r__l____h_____v____h___r____n____t___r____r__J_____  ___)
    __t___h_____r_t___n_____ _____ ___t_  ___  ____ _>__ _____ _T
    _v_  v__  _
_____  ___m_
_____  4____w  __,_
__y_,,,_____ __,__  2n__m  _,   ,,m__'_.,y?,_.___,,_,,_____,v__,?.___,  ,_
    .,x____,.  _.4,;___,,,,___.  ^,  __,_,_T5__,  __.__,_,_ny,,_,
_,  _ __,  ___y__?,  ,_'____ _?_  ____,_, m   ,   ,  _  ,  ,,      _,  __   ,
    .  ,   _  _  _n      .,,.,
_n%v____,_.  _  _4__,,.._.  __,,  __/
```

```
__,__ :4_ __  __ ?_.__?  _ 32m,__x J25__ _, _? ___  ____e,_ m___  __ ,__  _
      _v  ____;'_____       _
_x_4_  __,_ g_________                                  _'_5,
```

Output file 51: *GOCR – Image* i

```
___AE___m_e__a___9_____2_o____4_l_ o38____m  ____________________  _  __
    ____a_v___f__00__t___9_he_ten__t
 or__  f_án
f2stiy_ ,e+_en  _ _e ì:5                                        Q.
_e1  om l 261 m  av ^_astighet_en_Dxi_e 1:_5 u_t_g_ör sam_fäll ghet_en
G ång_gr_i f_ten _S _:l.
2
-   ------  '         '-^       '---'     .-_ _  _
.  t avta_  A 32g/2oo6 har de_ om ca 23 3oo m 2
o_x_ e l :_5_ _a vy_t t__a ts ._       _     _       '_
?
>
_
_

'    -   -- -   -  -      -        ^ ^  _  ' _  - 0___
```

Output file 52: *GOCR – Image* j

```
4___qJ_   u___
```

```
JTnnge (,sh,own  n  Appen_zi' AJ
_      h      c      d      ,
_,  T,ss,r_ct   16.97%o  35.59%o  16.72%o  36.7_%o  16.4%o
__  Ocr_d    2_.96%o  57.93%o  36.2%o  75.7_%o  51.51%o
_a  op,nocR  23.43%o  49.o1%o  17.33%o  7.442%o  17.6_%o
m
OCRupus'  32._7%o  15.79%o  37.64%o  41.52%o  39.3_%o
TOCR    4.74%o  4.413%o  13.22%o  7.2o7%o  11.26%o
Ahhyy     9.513%o  29.16%o  24.56%o  27.03%o  15.31%o

T_hl, l: ne,s1_lt,s oi th,e OCn ,scnn,s.

YU
```

```
_q    _U
__

_     4u                                    -
_
2U


M       h       C       d       ,
JTnnge (,sh,own   n   Appen_zi' AJ


T,s's',r_ct   l_  Ocr_d   Il_  Op,nOCR   GOCR
_ _ OCRupus' I _ TOCR   Ahhyy
```

Figur, 2: Jll1_,strnt on oi th,e ,scnn re,s1_lt,s.

Th, m,_n ,rrur v_lu, in incr,_s'ing urd,r fur ,_ch uf th, OCR tuuls' is' s' huwn
h,luw:

```
TocR      _.169%o
Abbyy     21.11%o
OpenOCR   22.9_%o
Tesseract    24.49%o
OCRopus   33.44%o
Ocrad      50.07%o
GocR     64.55%o
```

6

Output file 53: *GOCR – Image* k

```
__4___?v?v_____x?_0__?_?00___y_u?___t__c?__0_____________t?_?__J_______0_c______?
     _?_____t0______Jh?__lr/_r__t_______c0tf_r________f       _)     _0_        ___?
     _____
```

```
Testin_ OCR tools
Angelica Gahasio

1    E_glish
This is a pa6c _'ith a pict__rc to tcst thc OC_ tools.
Thc _agc cont_ins text in English _n_ S_'c_ish.
,c_?_',_ _ _a'__ _3, __ ___ ^ '-- :  ',  ',_',        ;_  ,  ,
_; n?_o ,Ji__?       _,                        0
0,  a_e_0?        ,,  ,            '  0        _
0,,__,',_,_  '

__^__,___  ' _                  _  '_?
```

```
_-_-____,';      0     c     0
_.,____J_____ _.                    _
_,____;'''''; ..
_______m.
__0,e,__,0,,?  '=,                        ,, 0
_?,  _?'
```

Fig__rc l: This is a c_t.

2 Swedish
Och så litc svcns_ techcn. Båtcn åhcr i sjön.
Dct _r hlött.

l

Output file 54: *GOCR – Image* l

```
_r_a___  _h___\_u_ ___e_ _e__u____ __\_ì_u_n
_____\ _^ ,_  _j_ __,h_i ___
___ ___
```

Output file 55: *GOCR – Image* m

```
Tl_c algoritl____ uscs a datal_asc of' lcttcrs a__d __u___crals to ___atcl_
    tl_c cl_ar-
actcrs. Tl_c algoritl____ will calculatc a ___atcl_i__g valuc f'or cacl_
    cl_aractcr a__d
sclccts tl_c o__tio__ witl_ tl_c l_cst valuc._1_
A wa_,T of' i_____rovi__g tl_c out__ut is to cl_cch if' tl_c sca____cd word
    cxists i__
a dictio__ar_,T. If' it docs__'t., it is lihcl_,T tl_at so___c of' tl_c
    a__al_,T_cd lcttcrs arc
wro__g., a__d so___c of' tl_c lcttcrs ___a_,T l_c cl_a__gcd so tl_c word
    ___atcl_cs a word i__
tl_c dictio__ar_,T. Tl_is is do__c l__,T addi__g a __c__alt_,T valuc to tl_c
    ___atcl_i__g valuc
if' tl_c a__al_,T_cd cl_aractcr rcsult i__ a word __ot i__ tl_c dictio__ar_,
    T. Tl_is ___ahcs
it __ossil_lc f'or a__otl_cr lcttcr to gct a l_cttcr ___atcl_i__g valuc if'
    tl_c word witl_
tl_c __cw lcttcr is __art of' tl_c dictio__ar_,T. Tl_c lcttcrs arc __ot
    alwa_,Ts cl_a__gcd.,
so___cti___cs a __o__-dictio__ar_,T word gcts a l_cttcr ___atcl_i__g scorc.,
     cvc__ witl_ tl_c
__c__alt_,T valuc., tl_a__ tl_c o__c witl_ tl_c cl_a__gcd lcttcrs tl_at is
    i__ tl_c dictio__ar_,T.
Evc__ if' usi__g a dictio__ar_,T ___a_,T l_cl__ tl_c out__ut accurac_,T it
    is __o guara__tcc
tl_at tl_c corrcct word will l_c f'ou__d._1_ A__ cxa_____lc of' tl_c usc of'
     a dictio__ar_,T
ca__ l_c scc__ i__ Figurc 1.
Otl_cr wa_,Ts to gct a l_cttcr out__ut is to usc ___acl_i__c lcar__i__g.,
    wl_crc tl_c
OCR tool ca__ l_c trai__cd to rccog__i_c ___orc la__guagcs or f'o__ts., or
    to usc tl_c
h__owlcdgc of' wl_icl_ cl_aractcr co___l_i__atio__s arc ___orc lihcl_,T to
    a____car.
```

```
Tl_c l_cttcr tl_c i___agc is., tl_c ___orc accuratc tl_c rcsult gcts.  If'
    tl_c OCR
algoritl_____ is a_____licd o__ a l_ori_o__tal tcxt witl_ vcr_,T littlc or __o
     __oisc a__d tl_c
lcttcrs arc wcll sc__aratcd f'ro___ cacl_ otl_cr it is __ossil_lc to gct a
    vcr_,T good
rcsult.  But if' tl_c i___agc is l_lurr_,T., war__cd or __ois_,T., or if'
    tl_c cl_aractcrs arc
___crgcd., it is lihcl_,T tl_at tl_c out__ut gcts worsc tl_a__ it would
    otl_crwisc._1., 2_
Tl_c out__ut accurac_,T will also dc__c__d o__ tl_c i___agc rcsolutio__.
OCR sof'twarc is gc__crall_,T l_cttcr at rccog__i_i__g ___acl_i__c writi__g
    tl_a__
l_a__dwriti__g. I__ ___acl_i__c writi__g tl_c sa___c cl_aractcr alwa_,Ts
    loohs tl_c sa___c.,
givc__ tl_at tl_c sa___c f'o__t is uscd. Witl_ l_a__dwriti__g tl_c
    cl_aractcrs di_ r a lot
dc__c__di__g o__ wl_o is writi__g., a__d cvc__ if' tl_c sa___c __crso__ l_as
     writtc__ a tcxt.,
tl_c sa___c cl_aractcr will l_c a littlc di_ rc__t cacl_ ti___c it is
    writtc__. _/1acl_i__c
writi__g of'tc__ co__sist of' ___orc disti__ct cl_aractcrs., a__d si__cc
    l_a__dwriti__g is
___orc varicd tl_a__ ___acl_i__c writi__g., it is l_ardcr f'or tl_c
    algoritl_____ to ___atcl_
tl_c cl_aractcrs to its datal_asc.
```

3

Output file 56: *GOCR – Image* n

B.5 OCRopus

```
i.e..
Ports
Fastigheten . Bilen . 4
Fr l re 19/8-09: Se azt ang Bilen 4 i lagf skJp
Del om 836 mg av fastigheten Bilen 4 egdr nybildning av
samfAlligheten Stapelbddden S:4 (se dverenskoelse i avyttrings
akt ang del om ca 3 020 mgav fastigheten Bilen 4)
Areal: 339 546 m2
Avst re 30/9-09: .Se ait ang Bilen 4 i lagf skJp
Del om 73im2 av fastigheten Bilen 4 utgdr Galeasen h
Del om 2 229 m2 av fastigheten Bilen 4 utgdr Galeasen k.
Areal: 336 580 m2
Forts fr5n fastigheten
Bilen 4
Fr l re 30/710: Se azt ang Bilen 4 i lagf skJp
ys VI
Sil 1
Del om 63 m2 av samftilligheten Flaggskepparen S:1 hartillagts
fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfAlligheten
Flaggskepparen s:1.
Areal: 332 894 m2
,Avst re 23/11-2010: Se akt ang Bilen 4 i lagf skJp
Del om 96 478 m2 av fastigheten Bilen 4 utgor fastigheten
Bilen 12.
Areal: 236 416 m2
Ys vl
```

Output file 57: *OCRopus – Image* a

```
-'y.,
PortS.
Fastigheten . i1n 4 e
Fr l re 19/8-09: Se azt ang Bilen 4 i lagf skJp
Del om 836 m2 av fastigheten Bilen 4 utgdr nybildning av
samfAlligheten Stapelbddden S:4 (se dverenskoelse i avyttrings-
akt ang del om ca 3 020 mgav fastigheten Bilen 4)
Areal: 339 546 m2
Awst re 30/909: e Se azt ang Bilen 4 i lagf skJp
Del om 731 sm2 av fastigheten Bilen 4 utgdr Galeasen L
lDel om 2 229 m2 av fastigheten Bilen 4 utgdr Galeasen 2.
Areal: 336 580 m2
Forts frar.astigheten
Bilen 4
Fr l re 30/710: Se azt ang Bilen 4 i 1agf skJp
ys Vt
Sif l
Del om 63 m2 av samfAlligheten Flaggskepparen S:1 har tillagts
fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfelligheten
Flaggskepparen s:1.
Areal: 332 894 d2
,Avst re 23/11-2010: Se akt ang Bilen 4 i lagf skap
Del om 96 478 m2 av fastigheten Bilen 4 utgor fastigheten
Bilen 12.
Areal: 236 416 m2
VG V1
```

Output file 58: *OCRopus – Image* b

```
portS .
renen-,v
Fastigheten i1ene 4 s
Fr l re 19/809: Se akt ang Bilen 4 i lagomp
a1 ds b5 m2 a,fastighten Silen 4 agdr nybildnin0 z
% samralligheten StaoNhadden S:4 (se overenskdgm1se i aortrings-
okt ang del om ca 3 020 m2 av fastigheten Bi1ef)
al: 339 546 m2
,Aat re 30/9-09; Se aM ang Bilen 4 i lagimp
61 om 737PA ay fastigheten Bilen 4 utgar sa0easen L
l01 om 2 229 m2 av fastijkten 8ilen 4 utyar saIeasen p
st[: 336 S80 m2
Forts fran fastigheten
Bilen 4
y6 yI
Sii 1
Fr l re 30/7-10: Se akt ang Bilen 4 i lagoep
el om as md av samalligheten Fla99skepparen S:1 bar til1a9s
fastigheten Bilen 4.
ai oe s Nz av taytisaetee stien 4 nar t1iiazts samsuiUkus
[Flaggskepparen s:L
aCl: 332 894 m2
AEt re 23/11-2010: Se akt ang Bilen 4 i lagiep
a1 ve 96 Os as raytIoneten stien 4 utner rasttsoten
Bilen 12.
S%(: 236 416 m2
yG VI
```

Output file 59: *OCRopus – Image* c

```
Enlttavtal A 515 och F 11 har del om ca 760 m2 av fastigheten
2009
e5ilen navyttrets,Se azt ang delar dom ca 2 . 060 e.resp 11 m2
av d efashighetem Bilen 7d.i slagf skap. Avst reg 16/3-2011
Enligt avtal A 34/2010 e har d del om ca 2 400 mg av fastigheten ee,
Bilen 4 avyttrats...Avst reg 15/7-10 - -
.Avst reg 15/7-1O: Se avytingsakt dang del om ca 2 400 m2 av
Bilen 4- -
1 Del om 2428 m2 av Bilen s4 sutgbr efastigheteaKlyvaren 1. A34/201
Del om 1 312 m24 av Bilen 4 utgor fastigheten Klyvaren g.
Areal: 332 840 o2
----FOrtS Sid )
Avst re 25/12011: Se akt ang Bilen 4 i lagf skap
Del om 1 017 m2 av Bilen 4 utgor Klippern 1.
Del om 813 mg av Bilen 4 utg6r Klippern d
'Del om 631 mg av Bilen 4 utgdr Klippern l
'Del om 1 08 m2 av Bilen 4 utgdr Klippern 4.
Del om 2 255 mmav Bilen 4 utgdr yosterbaten 1.
Del om 1 627 meav Bi len 4 utgdr Msterbaten 2.
1 Del om 1 784 mc av Bilen 4 utgor Koggen E
[Del om 1 666 m2 av Bilen 4 utgdr Koggen d
Areal: 225 545 m2
Avst reg 16/3-2011: Se azt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skJp
Del om 759 m2 av Bilen 4 utgor Bilen 13. A 515 och F 11
Areal: 224 786 m2
200 9
```

Output file 60: *OCRopus – Image* d

```
Faetigheten oxie 1:5
```

```
Pr l re 1 4-79: ee akt ang fastigh Oxie 1:1 m fl
Lel om 1O 801o7 3 av eamfalligh'ten Oxie e:2 har tillagte
5/1o7oY
faetigheten 0xie 1:5 (F st/;rtra/
4 dt/ Abod
re ,e 8ca.7 -d)
Enl'' avtaj d bar del av fastigheten oxie l:5 avybtrans,
1979
se akt ang. ca 75.00C vi av fastigheten Cxie 2l;2. Delen nar
vi,frgl. reg. l5lOl visAt oig inneholla 7.2O7.9 d.
j: 2 227 292,6 j
Ems. e7tol A 2Oto5l oaU del om i2.62,oc nan av CZie lt,, ave.
dh- bolaa faniEhnen por,amningen,av.crats..c, sh e'C
Avst Teg l7 /3884 se ak t ang Oxie l:l m fl
Uel oml0 225 me ag Oxiel:5 HtGUr FOtridmoinGen2
: 2 l8O 452 m''
```

Output file 61: *OCRopus – Image* e

```
Faiighetem Cxie l:5
Li/,--N
Bildad genom eammanlaggnimg aefaetigheterne Crie l:l,
Oxie lac ,, Oxie l9:l, Oxie 20:8, Oxie 25:9, Oxie 2S4 l och
Oxie 29:l.
'Areaieftersainld-nin:i 2.25l.222 3
Sdrskot ensip deol. reg. beslot den 54/4-92: eedrues049
Fr l. re .l7 t0-
i ee akt ang. Atg 244C m.f2. i Foeie
Pelar om 853C 3 av samfdldn omradet adj, l.956oC Pn av sam
fall5e omrdiet Chle s:l, 33.C7o 9 U' Rv etrdshgan . 28e och
-r.repon;n ev stadengam 7467,246O her tillagte faetighoten
orie l:5.
```

Output file 62: *OCRopus – Image* f

```
corts fran
Fastiheten 0sie l:5
Del om 1 261 m av fastigheten 0xie 1:5 utgUr samfalligheten
Ganggriften S:1.
Areal: 2 041 038 m2
Enligt stdlA3k9/206 hanelom ca23 d300 smg ave fastighetenses
c)
10xie 1:5 avyttrats.
```

Output file 63: *OCRopus – Image* g

```
corts fran
Fastiheten 0sie l:5
Del om 1 261 m av fastigheten 0xie 1:5 utgUr samfalligheten
Ganggriften S:1.
Areal: 2 041 038 m2
Enligt stdlA3k9/206 hanelom ca23 d300 smg ave fastighetenses
c)
10xie 1:5 avyttrats.
```

Output file 64: *OCRopus – Image* h

```
,i --a1
cm
,if No,v.
g-.
r
.
,
x
4 o
sed''oo,O re obe ees.0 asAse za4od sessoe e wseeisid 0 ses 0 ee,Ma soo - egy
    .o.
W
00
A
.
'd
L.s
, m
ee
e
t
$ o a 00
.n
e
E''
vo+
:
e
'Ze v
,
Y'6
ii
. ,
hvt
e e
r d
s
nAs
sy
s
.
i1J
o roo
00
e
Soc
i sA
serf'
P
Ps
en
e
4
e
00
f
s
e
-
4 i ./
=
-
,,
4
e
```

```
4 1
s
o
r-U
ats
ow
9
e
o
e
A
o
E
.6
isA
```

Output file 65: *OCRopus – Image* i

```
cortsfrah
Fastigheten 0sie I:5
Del om 1 261 m av fastigheten 0xie 1:5 utgor samfalligheten
Ganggriften S:1-
Areal: 2 041 038 m2
e e e
i
[EnligtavtalAR3I3006 hanalomca23 300 m2 av .ftighetenesses
j
108ies l.5 avyttrate,,,
Ins--
eth
AP---
seg
-i--
a
```

Output file 66: *OCRopus – Image* j

```
Image (shown m Appnds A7
a b c d e
gTact 16.97% 35.59% 16.72% 36.78% 16.49
D Ocrad 28.0% 57.93% 362% 75.78% 51.519
5 OpetCR 23.43% 4901% 17.33% 7.442% 17.68%
, GOCR 62 62% 41 04% 7768% 6286% 7853%
OCRops 32.87% 15.79% 3764%41.52% 39.38W
TOCR 474% 4413% 1322% 7207% 11269
Abbyy 9.513% 2916% 2466% 27.03% 15.31O
Table 1: Results of the OCR scans
ab
d e
Image (shown an Appnds A7
As Tessnao ls OOw tz OpeuOCR la GOCR
14 OCRopus if TOCR Abbyy
Figure 2: Illustraoon of the scan results
The meau ero value iu increasiug order for each of the OCR tools is showu
below
TOCR 81699
Abbyy 21.11A
OpenOCR 22989
Tesseract 24 499
OCRopus 33.449
Ocrad 50 07o
```

GOCR 64559

Output file 67: *OCRopus – Image* k

```
1English
Testing OClools
Angehca Gabas1o
This is a page with a picture to test the OCR tools
The page contains text in Enghsh and Swedish
2 Swedish
O s
es
m
f .- .
- M- N
'ure 1: This is a cat
Och sa hte sveuso teckeu. Bateu aker icou
Det ar blott
```

Output file 68: *OCRopus – Image* l

```
'pfica(Cheir.c cee.,orooo
(hN'thetrostAl
,
```

Output file 69: *OCRopus – Image* m

```
the algorithm uses a database of letters and numerals to match the char-
acters. The algorithm will calculate a matchmg value for each character and
selects the option with the best valuo[1
A wav of improving the ontot is to check if the scanned word exists in
a dictionarv. If It doesn t, It is likeb that some of the anabzed letters
    are
wrong, and some of the letters may be changed so the word matches a word m
the dictionary. This is done by adding a penaltv value to the matchmg value
if the analyzed character result m a word not m the dictionary. This makee
It possible for another letter to get a better matchmg value if the word
    with
the new letter is part of the dictionary. The letters are not always changed
sometimes a non-dictionary word gets a better matchmg score, even with the
penalr value, than the one with the changed letters that is m the dictionary
Even if usmg a dictionary mav help the ontot accuracy It is no guarantee
that the correct word will be found[1] An mample of the use of a dictionarv
can be seen in Figure 1
Other wavs to get a better output is to use machme learnmg, where the
OCR tool can be tramed to recognize more languages or fonts, or to use the
knowledge of which character combmations are more hiy to appear
The better the image is, the more accurate the result gets. If the OCR
algorithm is apphed on a horizontal text with very httle or no noise and the
letters are w1 separated from each other It is possible to get a verygood
result. But if the image is blurry, warped or noisv, or if the characters
    are
merged, it is likeb that the output gets worse than it would otherwiso[1, X
the ontot accuracy will also depend on the image resolntion
OCR software is generallv better at recognizing machme writing than
handwrIting. In machme wrIting the same character always looks the same
given that the same font is used. With handwriting the characters difler a
    lot
dependmg on who is writing, and even if the same person has written a tot
```

```
the same character will be a httle diflerent each time It is wrItten AIachme
wrItmg often consist of more distinct characters , and smce handwrItmg iS
more varied than machme writing , It is harder for the algorithm to match
the characters to Its database
```

Output file 70: *OCRopus – Image* n

B.6 TOCR

```
Forts 6.
Fastigheten Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av
samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-
akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
Areal : 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2
V G V!
Forts från fastigheten Sid 7
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m2 av samfälligheten Flaggskepparen S:1 har tillagts
fastigheten Bilen 4.
Del om 9 m av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:1.
Areal : 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal : 236 416 m2
V G V!
```

Output file 71: *TOCR – Image* a

```
Forts 6.
Fastigheten Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av
samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-
akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
Areal : 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2
Areal: 336 580 m2 '
V G V!
Forts från fastigheten Sid 7
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m2 av samfälligheten Flaggskepparen S:1 har tillagts
fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:1.
Areal : 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal : 236 416 m2
V G V!
```

Output file 72: *TOCR – Image* b

```
Forts 6.
Fastigheten Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
```

```
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av
samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings-
akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
Areal : 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2
V G Y!
Forts från fastigheten
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m2 av samfälligheten Flaggskepparen S:1 har tillagts
fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:1.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal : 236 416 m2
V G V!
```

Output file 73: *TOCR – Image* c

```
Enligt avtal A 515 och F 11 har del om ca 760 m2 av fastigheten
2009
Bilen 4 avyttrats. Se akt ang delar om ca 2 060 m2 resp 11 m2
av fastigheten Bilen 7 i lagf skåp. Avst reg 16/3-2011
Enligt avtal A 34/2010 har del om ca 2 400 m2 av fastigheten
Bilen 4 avyttrats. Avst reg 15/7-10
Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m av
Bilen 4.
Del om 2 428 m2 av Bilen 4 utgör fastigheten Klyvaren 1. A 34/201
Del om 1 312 m2 av Bilen 4 utgör fastigheten Klyvaren 2.
Areal : 332 840 m2
Forts sid 7
Avst reg 25/1-2011: Se akt ang Bilen 4 i lagf skåp
Del om 1 017 m2 av Bilen 4 utgör Klippern 1.
Del om 813 m2 av Bilen 4 utgör Klippern 2.
Del om 631 m2 av Bilen 4 utgör Klippern 3.
Del om 1 078 m2 av Bilen 4 utgör Klippern 4.
Del om 2 255 m2 av Bilen 4 utgör Kosterbåten 1.
2
Del om 1 627 m av Bilen 4 utgör Osterbåten 2.
Del om 1 784 m2 av Bilen 4 utgör Koggen 1.
Del om 1 666 m2 av Bilen 4 utgör Koggen 2.
Areal : 225 545 m2
Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skåp
Dei om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 11
Areal: 224 786 m2 2 0 0 9
```

Output file 74: *TOCR – Image* d

```
Fastigheten Oxie 1:5
Frgl reg 19/4-79: se akt ang fastigh Oxie 1:1 m fl
Del om 10 801,7 d av samfälligheten Oxie s:2 har tillagts
fastigheten Oxie 1:5 (F
A61, F60
Enl. avtal -7g7g har del av fastigheten Oxie 1:5 avyttrats,
```

```
se akt ang, ca 78.000 J av fastigheten Oxie 21;2. Delen har
vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 d.
Areal: 2 227 292,6
Enl. avtal A 20/1961 har del om 52.615,9 av 0xie 1:5, avs.
att bilda fastigheten Fornl ämningen 1 avyttrats. 7
Avst reg 17/3-88: se akt ang Oxie 1:1 m fl
Del om 14 225 m2 a oxie 1:5 utgör Fornlämningen 2
Areal: 2 180 452 m
```

Output file 75: *TOCR – Image* e

```
Fastigheten Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1,
Oxie 17:2, Oxie 19:1, Oxie 20:8, Oxie 25:9, Oxie 28:1 och
Oxie 29:1.
Areal efter sammanläggning: 2.231.222 d
Särskilt namn en1. reg. beslut den 15/4-92: Enedriksherg
Frgl. reg.l?/lO-.75 : se akt ang, stg 2440 m.fl. i Fosie
Delar om 8530 J av samfällda området add, 1.956,0 l av sam-
fällde området Oxie s:1, 33.507,2 av stedsägan.2834 och
22.12?,4 S av stedsägen 2467,2468 har tillagts fastigheten
Ori e 1: 5.
```

Output file 76: *TOCR – Image* f

```
Forts från
Fastigheten Oxie 1:5 4.
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal 2 041 038 m2
Eriligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
Oxie 1:5 avyttrats.
```

Output file 77: *TOCR – Image* g

```
Forts från
Fastigheten Oxie 1:5 4.
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal 2 041 038 m2
Eriligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
Oxie 1:5 avyttrats.
```

Output file 78: *TOCR – Image* h

```
i~otis ffån
Fastigheten 0#ie 1:5 4.
Del om 1 261 m av fastigheten Oxie 1:5 utgör samf811igheten
Gånggritten S:1.
nìigt avtal A 3?9/2006 har del om ca 23 300 m2 av fastigheten
Oxie 1:5 avyttrats
```

Output file 79: *TOCR – Image* i

```
Forts från
Fastigheten Oxie 1:5 4.
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal 2 041 038 m2
Eriligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
Oxie 1:5 avyttrats.
```

Output file 80: *TOCR – Image* j

```
Image (shornn in Appendix A)
a b c d e
Tesseract 16.97% 35.59% 16.72% 36.78% 16.4%
$ Ocrad 28.96% 57.93% 36.2% 75.78% 51.51%
OpenOCR. 23.43% 49.01% 17.33% 7.442% 17.68%
GOCR. 62.62% 41.04% 77.68% 62.86% 78.53%
OCR.opus 32.87% 15.79% 37.64% 41.52% 39.38%
TOCR. 4.74% 4.413% 13.22% 7.207% 11.26%
Abbyy 9.513% 29.16% 24.56% 27.03% 15.31%
Table 1: Results of the OCR scans.
80
60
40
20
0
b c d e
Image (shornn in Appendix A)
Tesseract Ocrad OpenOCR. GOCR.
OCR.opus TOCR. Abbyy
Figure 2: Illustration of the scan results.
The mean error value in increasing order for each of the OCR. tools is shown
below:
TOCR 8.169%
Abbyy 21.11%
OpenOCR 22.98%
Tesseract 24.49%
OCRopus 33.44%
Ocrad 50.07%
GOCR 64.55%
6
```

Output file 81: *TOCR – Image* k

```
Testing OCR tools
Angelica Gabasio
1 English
This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.
Figure 1: This is a cat.
2 Swedish
Och så lite svenska tecken. Båten åker i sjön.
Det är blött.
1
```

Output file 82: *TOCR – Image* l

```
p cd har c acc nho
ANGEL A ABA5 c
```

Output file 83: *TOCR – Image* m

The algorithm uses a clatabase of letters ancl numerals to match the characters. The algorithm will calculate a matching value for each character ancl
selects the option with the best value.[1]
A way of improving the output is to check if the scannecl worcl exists in a clictionary. If it cloesn't, it is likely that some of the analyzecl letters are
wrong, ancl some of the letters may be changecl so the worcl matches a worcl in
the clictionary. This is clone by aclcling a penalty value to the matching value
if the analyzecl character result in a worcl not in the clictionary. This makes
it possible for another letter to get a better matching value if the worcl with
the new letter is part of the clictionary. The letters are not always changecl,
sometimes a non-clictionary worcl gets a better matching score, even with the
penalty value, than the one with the changecl letters that is in the clictionary.
Even if using a clictionary may help the output accuracy it is no guarantee that the correct worcl will be founcl.[1] An example of the use of a clictionary
can be seen in Figure 1.
Other ways to get a better output is to use machine learning, where the OCR tool can be trainecl to recognize more languages or fonts, or to use the knowleclge of which character combinations are more likely to appear.
The better the image is, the more accurate the result gets. If the OCR algorithm is appliecl on a horizontal text with very little or no noise ancl the
letters are well separate<l from each other it is possible to get a very goocl
result. But if the image is blurry, warpecl or noisy, or if the characters are
mergecl, it is likely that the output gets worse than it woukl otherwise.[1, 2]
The output accuracy will also clepencl on the image resolution.
OCR software is generally better at recognizing machine writing than hanclwriting. In machine writing the same character always looks the same, given that the same font is usecl. With hanclwriting the characters cliffer a lot
clepencling on who is writing, ancl even if the same person has written a text,
the same character will be a little clifferent each time it is written. Machine
writing often consist of more clistinct characters, ancl since hanclwriting is
more variecl than machine writing, it is harcler for the algorithm to match the characters to its clatabase.
3

Output file 84: *TOCR – Image* n

B.7 Abbyy

```
  Forts      (          ' !"X 6'
Fastigheten; Bilen 4 N_
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfäll igheten Stapel bädden S:4 (se överenskommelse i avyttrings-
2
akt ang  del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
2
Del om 2 229 m  av fastigheten Bilen 4 utgör Galeasen 2~
Areal: 336 580 m2
V G V!
Forts   från fastigheten bia '
Bilen 4_
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m   av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfäll igheten
Flaggskepparen s:l.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 85: *Abbyy – Image* a

```
  Forts      (   s '
Fastigheten Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
2
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfäll igheten Stapel bädden S:4 (se överenskommelse i avyttrings-
2
akt ang   del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
2
Del om 2 229 m  av fastigheten Bilen 4 utgör Galeasen 2~
Areal: 336 580 m2
V G V!
Forts  från fastigheten bia 1
Bilen 4_
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
2
Del om 63 m   av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfäll igheten
Flaggskepparen s:l.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m  av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
```

V G V!

Output file 86: *Abbyy – Image* b

```
   Fastigheten     Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
2
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfäll igheten Stapel bädden S:4 (se överenskommelse i avyttrings-
2
akt ang   del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
/st reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
il om 737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
2
:1 om 2 229 nr av fastigheten Bilen 4 utgör Galeasen 2~
-eal: 336 580 m2
V G V!
Forts  från fastigheten                                        Sld 7
Bilen 4_
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
2
Del om 63 m   av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfäll igheten
Fl aggskepparen s:l.
Areal: 332 894 m2
vst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 87: *Abbyy – Image* c

```
  Enligt   avtal A 515 och F 11   har: del om ca 760 m2 av fastigheten
2009__ _
Bilen 4 avyttrats.   Se akt ang delar om ca 2 060 m2resp 11 m2
av fastigheten Bilen 7 i lagf skåp. Avst reg 16/3-2011__
Enligt avtal A 34/2010   har del om ca 2 400 m2 av fastigheten
Bilen 4 avyttrats.   Avst reg 15/7-10
Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m2 av
Bilen 4.
Del om 2 428 m2 av Bilen 4 utgör fastigheten Klyvaren 1. A 34/201
Del om 1 312 m2 av Bilen 4 utgör fastigheten Klyvaren 2._______
Areal: 332 840 m2_____________________________
------------------------------------------------------------ Forts sid 7
Avst reg 25/1-2011: Se akt ang Bilen 4 i lagf skåp
! Del om 1 017 m2 av Bilen 4 utgör Klippern 1.
I Del om 813 m2 av Bilen 4 utgör Klippern 2.
[ Del om 631 m2 av Bilen 4 utgör Klippern 3.
i Del om 1 078 m2 av Bilen 4 utgör Klippern 4.
Del om 2 255 m2 av Bilen 4 utgör   ;<osterbåten 1.
i 2
    Del om 1 627 m av Bilen 4 utgör   JOsterbåten 2.
f Del om 1 784 m2 av Bilen 4 utgör Koggen 1.
I Del om 1 666 m2 av Bilen 4 utgör Koggen 2.
Areal: 225 545 m2
I
Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
```

```
fastigheten Bilen 7 i lagf skåp
De? om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 11
Areal: 224 786 m2 2 0 0 9
```

Output file 88: *Abbyy – Image* d

```
   Fastigheten     Oxie 1:5
Frgl reg 19/4-79: se akt ang fastigh Oxie 1:1 a fl
Del om 10 801,7 rfi av samfälligheten Oxie s:2 har tillagts
fastigheten Oxie 1:5 (| f^/j^)
    , . A61, F60
Enl. avtal  har  del av fastigheten Oxie 1:5 avyttrats,
se akt ang.   ca 78.000 nf av fastigheten Oxie 21; 2. Delen har
vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 ma.
Areal: 2 22 7 2 92,6 af
Enl. avtal A 20/1981 har del om 3 2,615,9 m2 av Oxie 1:5, avs.
att bilda fastigheten Porn1 amningen 1 avyttrats. %-*s)
Avst  reg I7/3-88:  se akt  ang Oxie  1:1 m fl
Del  om  1^   225 m2  av Oxie   1:5  utgör  Fornlämningen 2
Areal:   2   180  ^52 m2
```

Output file 89: *Abbyy – Image* e

```
     57, c
Fastigheten Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1,
Oxie 17:2, Oxie 19:1, Oxie 20:8,  Oxie 25:9, Oxie 28:1 och
Oxie 29:1.
.Areal efter sammanläggning:  2.231.222 nf
Särskilt namn enl. reg. beslut den 15/4-92: i        I ksb.erc}
Frgl. reg.17/10-75     : se akt ang. stg 2440 m.fl. i Fosie
Delar om 85,0 m av samfällda området adj,  1.956,0 nf av sam-
fällda området Oxie s:l, 33.507,2 nf av stadsägan. 2834 och
22.122,4 nf ?.v stadsägan 2^67, 2468 har tillagts fastigheten
Oxie 1:5.
```

Output file 90: *Abbyy – Image* f

```
  Forts   från
Fastigheten Oxie 1:5
4,
Del om 1 261 m jiv^fastigheten Oxie 1:5 utgör samfäll i gheten
Gång^riften S:l.
Areal: 2 041 038 m2_____________________________________
Enligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
Ox i e 1 :5 avyttrats._____ . ______\_________________
```

Output file 91: *Abbyy – Image* g

```
  Forts   från
Fastigheten Oxie 1:5
4,
Del om 1 261 m jiv^fastigheten Oxie 1:5 utgör samfäll i gheten
Gång^riften S:l.
Areal: 2 041 038 m2_____________________________________
Enligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
```

```
Ox i e 1 :5 avyttrats._____ . _______\_________________
```

Output file 92: *Abbyy – Image* h

Output file 93: *Abbyy – Image* i

```
  Forts    från
Fastigheten Oxie 1:5
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfäll igheten
Gånggriften S:l.
Areal: 2 041 038 m2 m
Enligt avtal A 329/2006 har del om ca 23 300 m av fastigheten
Oxie 1:5 avyttrats.___ ____1_ r> __
```

Output file 94: *Abbyy – Image* j

```
                Image (shown in A}                pendix A)
        a       b       c       d       e
Tesseract       16.97%  35.59%  16.72%  36.78%  16.4%
Ocrad   28.96%  57.93%  36.2%   75.78%  51.51%
OpenOCR 23.43%  49.01%  17.33%  7.442%  17.68%
GOCR    62.62%  41.04%  77.68%  62.86%  78.53%
OCRopus 32.87%  15.79%  37.64%  41.52%  39.38%
TOCR    4.74%   4.413%  13.22%  7.207%  11.26%
Abbyy   9.513%  29.16%  24.56%  27.03%  15.31%
Table 1: Results of the OCR seans.
Til
1
b c d
Image (shown in Appendix A)
ils is shown
III Tesseract 11 Ocrad III OpenOCR GOCR
11 OCRopus ll TOCR Abbyy
Figure 2: Illustration of the sean results.
TOCR 8.169%
Abbyy 21.11%
OpenOCR 22.98%
Tesseract 24.49%
OCRopus 33.44%
Ocrad 50.07%
GOCR 64.55%
f)
```

Output file 95: *Abbyy – Image* k

```
  Testing   OCR tools
Angelica Gabasio
1 English
This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.
W
II
Figure 1: This is a cat.
2 Swedish
Och så lite svenska, tecken. Båten åker i sjön.
```

```
Det är blött.
i
```

Output file 96: *Abbyy – Image* l

```
   optical       CWarae^e-r rtcoanvilon
2013
```

Output file 97: *Abbyy – Image* m

```
  The  algorithm uses a database of letters and numerals to match the char-
acters. The algorithm will calculate a matching value for each character and
selects the option with the best value. [1]
A way of improving the output is to check if the scanned word exists in
a dictionary. If it doesn't, it is likely that some of the analyzed letters
    arc
wrong, and some of the letters may be changed so the word matches a word in
the dictionary. This is done by adding a penalty value to the matching value
if the analyzed character result in a word not in the dictionary. This makes
it possible for another letter to get a better matching value if the word
    with
the new letter is part of the dictionary. The letters arc not always changed
    ,
sometimes a non-dictionary word gets a better matching score, even with the
penalty value, than the one with the changed letters that is in the
    dictionary.
Even if using a dictionary may help the output accuracy it is no guarantee
that the correct word will be found. [1] An example of the use of a
    dictionary
can be seen in Figure 1.
Other ways to get a better output is to use machine learning, where the
OCR tool can be trained to recognize more languages or fonts, or to use the
knowledge of which character combinations arc more likely to appear.
The better the image is, the more accurate the result gets. If the OCR
algorithm is applied on a horizontal text with very little or no noise and
    the
letters arc well separated from each other it is possible to get a very good
result. But if the image is blurry, warped or noisy, or if the characters
    arc
merged, it is likely that the output gets worse than it would otherwise.[1,
    2]
The output accuracy will also depend on the image resolution.
OCR software is generally better at recognizing machine writing than
handwriting. In machine writing the same character always looks the same,
given that the same font is used. With handwriting the characters differ a
    lot
depending on who is writing, and even if the same person has written a text,
the same character will be a little different each time it is written.
    Machine
writing often consist of more distinct characters, and since handwriting is
more varied than machine writing, it is harder for the algorithm to match
the characters to its database.
```

Output file 98: *Abbyy – Image* n

B.8 Leadtools

```
64
Forts i
Fastigheter Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av
1
samfälligheten Stapelbädden S:4 (se överenskommelse i avyttrings- ,
akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
T Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737m2 av fastigheten Bilen 4 utgör Galeasen 1.
i 1Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen
Areal: 336 580 m2
. V G V!
Förts från fastigheten Sid 7
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m2 av samfälligheten Flaggskepparen S:1 har tillagts
fastigheten Bilen 4.
Del om 9 m 2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s: 1.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 99: *Leadtools – Image* a

```
  6  ,
Forts ^
Fastigheten Bilen 4
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
Del om 836 m2 av fastigheten Bilen 4 utgör nybildning av
I
samfälligheten Stapelbädden 5:4 (se överenskommelse i avyttrings- ,
! 1
akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737m2 av fastigheten Bilen 4 utgör Galeasen le
i 1  Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2
V G V!
Forts från fastigheten Sid i
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m2 av samfälligheten Flaggskepparen 5:1 har tillagts
fastigheten Bilen 4.
Del om 9 m2 av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s: 1.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 100: *Leadtools – Image* b

```
  Forts
Fastigheten Bilen 4
64
..._
fulseg 19/8-09: Se akt ang Bilen 4 i Lagt' skåp
Del om 836 m 7 av fastigheten Bilen 4 utgör nybildning av
,
,
samfälligheten Stapelbädden S:4 (se överenskommelsei avyttrings- ,
i
i akt ang del om ca 3 020 m2 av fastigheten Bilen 4)
Areal: 339 546 m2
‘,52..s1,22,2219-09: Se akt ang Bilen 4 i lagf skåp
/
Del om737 m2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2.,
Areal: 336 580 m2
V G V!
Forts från fastigheten Sid 7
Bilen 4
Fr i re 30/7-10: Se akt ang Bilen 4 i lagt' skåp
Del om 63 m2 av samfälligheten Flaggskepparen 5:1 har tillagts
fastigheten Bilen 4.
D i el om 9 m2av fastigheten Bilen 4 har tillagts samfälligheten
J
naggskepparen 5:1.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i hyr skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 101: *Leadtools – Image* c

```
  Enligt   avtal A 515 och F 11 har del om ca 760 m2av fastigheten
2009
Bilen 4 avyttrats. Se akt ang delar omca 2 060} resp 11 m2 _
av fastigheten Bilen 7 i lagf skåp. Avst reg 16/3-2011
Enligt avtal A 34/2010 har del om ca 2 400 m2 av fastigheten
I
_Bilen 4 avyttrats . Avst reg 15/7710
Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m2 av
Bilen 40
1  Del om? 428 m2 av Bilen 4 utgör fastigheten Klyvaren 1. A 34/201
--  I Del om 1 312 m2 av Bilen 4 utgör_fastigheten KlYvaren 24
t
Areal: 33? 840 m2
Fortssid7
Avst reg 25/1-2011: Se akt ang Bilen 4 i lagf skåp
Del om1 017 m2 av Bilen 4 utgör Klippern 1.
Del om 813 m2av Bilen 4 utgör Klippern 2.
Del om 631 m2av Bilen 4 utgör Klippern 3.
Del om1 078 m2 av Bilen 4 utgör Klippern 4.
Del om 2 255 m2av Bilen 4 utgör Kosterbåten 1.
2
Del om1 627 m av Bilen 4 utgör K)sterbåten 2.
Del om 1 784 m2 av Bilen 4 utgör Koggen 1.
Del om 1 666 m2 av Bilen 4 utgör Koggen 2.
Areal: 225 545 m2
Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skåp
1 Del om 759 m2av Bilen 4 utgör Bilen 13. A 515 och F 11
```

```
Areal: 224 786m2
2 0 09
```

Output file 102: *Leadtools – Image* d

```
   Fastigheten     Oxie 1:5
Frgl reg 19/4-79: se akt ang fastigh Oxie 1:1 m fl
av samfälligheten Oxie s:2 har tillagts
Del om 10 801 97
d
fastigheten Oxie 1:5
(y Wiffl)
590r ,o,cpc
A61, F60
Enl. avtal har del av fastigheten Oxie 1:5 avyttrats,
1979
se akt ang. ca 78.000 å av fastigheten Oxie 21;2. Delen har
vid frgl. reg. 15/1-81 visat sig innehålla 7.2879 1112.
Areal: 2 227 292,6 ni
Enl. avtal A 20/1981 har del om 32.615,9 å av Oxie 1:5, avs.
att bilda fastigheten Fornlämningen 1 avyttrats.6' 7 '21 c4'sv)
Avst reg 17/3-88: se akt ang Oxie 1:1 m fl
Del om 14 225 m 2 ax Oxie 1:5 utgör Fornlämningen 2
452 m'
2 180
Areal:
```

Output file 103: *Leadtools – Image* e

```
  4
Fastigheten Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1,
Oxie 17:2, Oxie 19:1, Oxie 20:8, Oxie 25:9, Oxie 28:1 och
Oxie
,AL25.2.22122=haailW 2.231.222 af
Särskilt namn en1 teg. beslut den 15/4-92: Uzedriksherg_
a : m 6 s 1Fx. e.1710--.12 a.: se akt ang. stg 2440 m.fl. i Fosie
av samfälidg området adj, 1.956,0 Ä av sams
Delar om 85,0
J
fällde området Oxie s:1 ? 55.507,2 rlf av stadsägan.2854 och
92.122,4 d sv stadsägan 2467,2468 har tillagts fastigheten
Oxie 1:5.
```

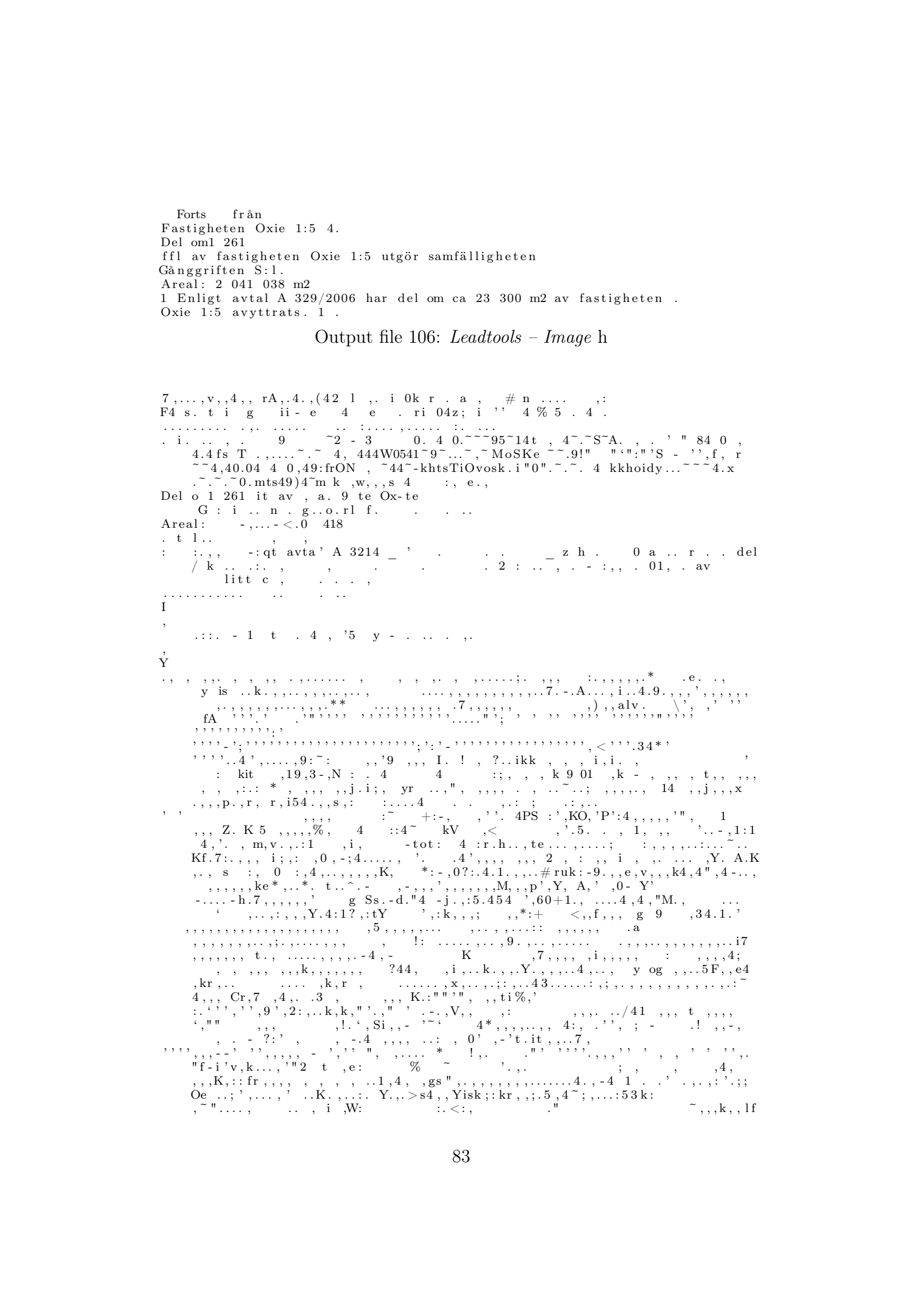
Output file 104: *Leadtools – Image* f

```
  Forts    från
Fastigheten Oxie i:5 40
Del om 1 261
ffi
av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:l.
Areal: 2 041 038 m2
1 Enligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
.
Oxie 1:5 avyttrats. __ .
_e__
```

Output file 105: *Leadtools – Image* g

Forts från
Fastigheten Oxie 1:5 4.
Del om1 261
ffl av fastigheten Oxie 1:5 utgör samfälligheten
Gånggriften S:1.
Areal: 2 041 038 m2
1 Enligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten .
Oxie 1:5 avyttrats. 1 .

Output file 106: *Leadtools – Image* h



Output file 107: *Leadtools – Image* i

```
  Forts    från
Fastigheten Oxie 1:5 4.
I Del om1 261 m2 av fastigheten Oxie 1:5 utgör samfälligheten
Gånggri ften S:1.
Areal: 2 041 038 m2

'Ungt avtal A 329/2006 har del om ca 23 30 m2 av fo, - igheten .
0 4
Oxie 1:5 avyttrats. An, .
    ?,
8 1
." 9
1
1
- - - -
e
. ,
-
Ta,
/1  e
a1  ,
O
```

Output file 108: *Leadtools – Image* j

```
Image (shown in Appendix A)
a
```

```
Tesseract 16.97% 35.59% 16.72% 36.78% 16.4%
Ocrad 28.96% 57.93% 36.2% 75.78% 51.51%
OpenOCR 23.43% 49.01% 17.33% 7.442% 17.68%
GOCR 62.62% 41.04% 77.68% 62.86% 78.53%
OCRopus 32.87% 15.79% 37.64% 41.52% 39.38%
TOCR 4.74% 4.413% 13.22% 7.207% 11.26%
Abbyy 9.513% 29.16% 24.56% 27.03% 15.31%
Table 1: Results of the OCR scans.
80
fl
(1) 60
z:s.) 1111!  I
(1)
(1) gt. 40 I 11 1  I
LI
20-
I
0
a
Image (shown in Appendix A)
I I Tesseract Ii Ocrad I  I OpenOCR GOCR
II OCRopus II TOCR I I Abbyy
Figure 2: Illustration of the scan results.
The mean error value in increasing order for each of the OCR tools is shown
below:
TOCR 8.169%
Abbyy 21.11%
OpenOCR 22.98%
Tesseract 24.49%
OCRopus 33.44%
Ocrad 50.07%
GOCR 64.55%
6
```

Output file 109: *Leadtools – Image* k

```
  Testing  OCR tools
Angelica Gabasio
1 English
This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.
il , fra
. \
i ' 4
"niors
Eigure 1: This is a cat.
2 Swedish
Och så lite svenska tecken. Båten åker i sjön.
Det är blött.
1
```

Output file 110: *Leadtools – Image* l

```
 a  P k. i cak- C bara:åar. re_co n g 'Uön
L.;
ANGEL\ CA GIABN5 0
2.013
```

Output file 111: *Leadtools – Image* m

```
The algorithm uses a database of letters and numerals to match the char-
acters. The algorithm will calculate a matching value for each character and
selects the option with the best value. [1]
A way of improving the output is to check if the scanned word exists in
a dictionary. If it doesn't, it is likely that some of the analyzed letters
    are
wrong, and some of the letters may be changed so the word matches a word in
the dictionary. This is done by adding a penalty value to the matching value
if the analyzed character result in a word not in the dictionary. This makes
it possible for another letter to get a better matching value if the word
    with
the new letter is part of the dictionary. The letters are not always changed
    ,
sometimes a non-dictionary word gets a better matching score, even with the
penalty value, than the one with the changed letters that is in the
    dictionary.
Even if using a dictionary may help the output accuracy it is no guarantee
that the correct word will be found. [1] An example of the use of a
    dictionary
can be seen in Figure 1.
Other ways to get a better output is to use machine learning, where the
OCR tool can be trained to recognize more languages or fonts, or to use the
knowledge of which character combinations are more likely to appear.
The better the image is, the more accurate the result gets. If the OCR
algorithm is applied on a horizontal text with very little or no noise and
    the
letters are well separated from each other it is possible to get a very good
result. But if the image is blurry, warped or noisy, or if the characters
    are
merged, it is likely that the output gets worse than it would otherwise.[1,
    2]
The output accuracy will also depend on the image resolution.
OCR software is generally better at recognizing machine writing than
handwriting. In machine writing the same character always looks the same,
given that the same font is used. With handwriting the characters differ a
    lot
depending on who is writing, and even if the same person has written a text,
the same character will be a little different each time it is written.
    Machine
writing often consist of more distinct characters, and since handwriting is
more varied than machine writing, it is harder for the algorithm to match
the characters to its database.
3
```

Output file 112: *Leadtools – Image* n

B.9 OCR API Service

```
  Frgl   reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
2
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfälligheten Stapel bädden S:4 (se överenskommelse i avyttrings-
2
akt ang   del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 ay fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2.
Areal: 336 580 m2
V G V!
Forts   från fastigheten 51d 7
Bilen 4__
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
2
Del om 63 m  av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:l.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 113: *OCR API Service – Image* a

```
  Forts   s
Fastigheterf Bilen 4 \)
6,
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skap
2
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfäll igheten Stapel bädden S:4 (se överenskommelse i avyttrings
2
akt ang   del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 rr 2 av fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2
Areal: 336 580 m2
V G V!
Forts  från fastigheten 2)1 a 1
Bilen 4_
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m   av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfäll igheten
Flaggskepparen s:l.
Areal: 332 894 m2
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
o
Del om 96 478 m  av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
```

V G V!

Output file 114: *OCR API Service – Image* b

```
  Forts
Fastigheten Bilen 4
6,
Frgl reg 19/8-09: Se akt ang Bilen 4 i lagf skåp
2
Del om 836 m  av fastigheten Bilen 4 utgör nybildning av
samfälligheten Stapel bädden S:4 (se överenskommelse i av
2
akt ang   del om ca 3 020 m  av fastigheten Bilen 4)
Areal: 339 546 m2
Avst reg 30/9-09: Se akt ang Bilen 4 i lagf skåp
Del om 737 m2 ay fastigheten Bilen 4 utgör Galeasen 1.
Del om 2 229 m2 av fastigheten Bilen 4 utgör Galeasen 2
Areal: 336 580 m2
V G V!
Forts  från fastigheten Sld 7
Bilen 4
Frgl reg 30/7-10: Se akt ang Bilen 4 i lagf skåp
Del om 63 m   av samfälligheten Flaggskepparen S:l har tillagts
fastigheten Bilen 4.
2
Del om 9 m av fastigheten Bilen 4 har tillagts samfälligheten
Flaggskepparen s:l.
Areal: 332 894 m2
t
Avst reg 23/11-2010: Se akt ang Bilen 4 i lagf skåp
Del om 96 478 m2 av fastigheten Bilen 4 utgör fastigheten
Bilen 12.
Areal: 236 416 m2
V G V!
```

Output file 115: *OCR API Service – Image* c

```
  Enligt   avtal A 515 och F 11   har del om ca 760 nr av fastigheten
_______ 2009______ "
Bilen 4 avyttrats.   Se akt ang delar om ca 2 060 m2 resp H m2
av fastigheten Bilen 7 i lagf skåp. Avst reg 16/3-2011
Enligt avtal A 34/2010   har del om ca 2 400 nr av fastigheten
Bilen 4 avyttrats.   Avst reg 15/7-10_____________________
Avst reg 15/7-10: Se avyttringsakt ang del om ca 2 400 m2 av
Bilen 4._____________________________ _ _______
Del om 2 428 m2 av Bilen 4 utgör fastigheten Klyvaren 1. A 34/201
Del om 1 312 m2 av Bilen 4 utgör fastigheten Klyvaren 2.__
Areal: 332 840 m2_____________________________
-------------------------------------------------------------_____ Forts
     sid 7
Avst reg25/l-2011: Se akt ang Bilen 4 i lagf skåp_
I Del om 1 017 m  av Bilen 4 utgör Klippern 1.
I 2
I Del om 813 m  av Bilen 4 utgör Klippern 2.
! Del om 631 m  av Bilen 4 utgör Klippern 3.
i Del om 1 078 m2 av Bilen 4 utgör Klippern 4.
f 2
i Del om 2 255 m  av Bilen 4 utgör   Kosterbåten 1.
I Del om 1 627 m av Bilen 4 utgör   Kosterbåten 2.
Del om 1 784 m2 av Bilen 4 utgör Koggen 1.
Del om 1 666 m2 av Bilen 4 utgör Koggen 2.
```

```
Areal: 225 545 m2
Avst reg 16/3-2011: Se akt ang delar om ca 2060 m2 resp 11 m2 av
fastigheten Bilen 7 i lagf skåp
I De? om 759 m2 av Bilen 4 utgör Bilen 13. A 515 och F 11
Areal: 224 786 m2 0 n n a
```

Output file 116: *OCR API Service – Image* d

```
   Fastigheten     Oxie 1:5
Frgl reg 19/4-79: se akt ang fastigh Oxie 1:1 m fl
Del om 10 801,7 nf av samfälligheten Oxie s:2 har tillagts
fastigheten Oxie 1:5 (j
^ ^ 3^ S~#Ö. S- sr.
Enl. avtal ^^g^g^ ^ar öel av fastigheten Oxie 1:5 avyttrats,
se akt ang.   ca 78.000 nf av fastigheten Oxie 21; 2. Delen har
vid frgl. reg. 15/1-81 visat sig innehålla 7.287,9 m.
Areal: 2 227 292,6 rf
Enl. avtal A 2Q/1981 har del om 3 2,615,9 m* av Oxie 1:5, avs.
att bilda fastigheten Pornl amningen 1 avyttrats.
Avst reg 17/3-88: se akt ang Oxie 1:1 m fl
Del om Ik 225 2 av Oxie 1:5 utgör Fornlämningen 2
:  2 180 k$2 m2
```

Output file 117: *OCR API Service – Image* e

```
   Fastigheten     Oxie 1:5
Bildad genom sammanläggning av fastigheterna Oxie 1:1,
Oxie 17:2, Oxie 19:1, Oxie 20:8, Oxie 25:9, Oxie 28:1 och
Oxie 29:1.
Areal efter sammanläggning: 2.231.222 m
Särskilt namn enl. reg. beslut den .15/4-92:.. Fxe.driksb.erg^
Frgl. reg, 17/10-75     : se akt ang. stg 2440 m.fl. i Fosie
Delar cm 85,0 af av samfällda området adj, 1,956,0 ra2 av sam-
fällda området Oxie s:l, 33.507,2 i av stadsägan. 2834 och
22,122,4 i av stadsägan 2467,2468 har tillagts fastigheten
Oxie 1:5.
```

Output file 118: *OCR API Service – Image* f

```
  Forts  från
Fastigheten Oxie 1:5
Del qm_l_261 m  av fastigheten
Gång^riften S:l.
Areal: 2 041 038 m
Enl igt avtal A 329/2006 har del
Oxie 1:5 avyttrats.____
4.
3xiel:5 utgör samfäll igheten
2
om ca 23 300 m  av fastigheten
```

Output file 119: *OCR API Service – Image* g

```
  Forts  från
Fastigheten Oxie 1:5
4.
```

```
Del om 1 261 m av fastigheten Oxie 1:5 utgör samfäl l igheten
Gång^riften S:l.
Areal: 2 041 038 m2
Enligt avtal A 329/2006 har del om ca 23 300 m2 av fastigheten
Oxie 1 ^^y^yttrats^__________j__________
```

Output file 120: *OCR API Service – Image* h

Output file 121: *OCR API Service – Image* i

```
  Forts    från
Fastigheten Oxie 1:5
Del om 1 261 m   av fastigheten Oxie 1:5 jJtgör samfälligheten
Gänggriften S:1
Areal: 2 041 038 m
2
4
1 Enligt avtal A 329/2006 har del om ca 23 300 m2 av f
Oxie 1:5 avvttrats.
```

Output file 122: *OCR API Service – Image* j

```
  Image   (shown in Appendix A)
        a       b       c       d       e
Tesseract       16.97%  35.59%  16.72%  36.78%  16.4%
Ocrad   28.96%  57.93%  36.2%   75.78%  51.51%
OpenOCR 23.43%  49.01%  17.33%  7.442%  17.68%
GOCR    62.62%  41.04%  77.68%  62.86%  78.53%
OCRopus 32.87%  15.79%  37.64%  41.52%  39.38%
TOCR    4.74%   4.413%  13.22%  7.207%  11.26%
Abbyy   9.513%  29.16%  24.56%  27.03%  15.31%
Table 1: Results of the OCR seans.
Il Tesseract 11 Ocrad 11 OpenOCR GOCR
Il OCRopus II TOCR Abbyy
Figure 2: Illustration of the scan results.
The mean error value in increasing order for each of the OCR tools is shown
below:
TOCR    8.169%
Abbyy   21.11%
OpenOCR 22.98%
Tesseract       24.49%
OCRopus 33.44%
Ocrad   50.07%
GOCR    64.55%
6
```

Output file 123: *OCR API Service – Image* k

```
  Testing   OCR tools
Angelica Gabasio
1 English
This is a page with a picture to test the OCR tools.
The page contains text in English and Swedish.
Figure 1: This is a cat
2 Swedish
```

```
Och så lite svenska tecken. Båten åker i sjön.
Det är blött.
1
```

Output file 124: *OCR API Service – Image* l

Output file 125: *OCR API Service – Image* m

```
  The  algorithm uses a database of letters and numerals to match the char-
acters. The algorithm will calculate a matching value for each character and
selects the option with the best value. [1]
A way of improving the output is to check if the scanned word exists in
a dictionary. If it doesn't, it is likely that some of the analyzed letters
    are
wrong, and some of the letters may be changed so the word matches a word in
the dictionary. This is done by adding a penalty value to the matching value
if the analyzed character result in a word not in the dictionary. This makes
it possible for another letter to get a better matching value if the word
    with
the new letter is part of the dictionary. The letters are not always changed
    ,
sometimes a non-dictionary word gets a better matching score, even with the
penalty value, than the one with the changed letters that is in the
    dictionary.
Even if using a dictionary may help the output accuracy it is no guarantee
that the correct word will be found. [1] An example of the use of a
    dictionary
can be seen in Figure 1.
Other ways to get a better output is to use machine learning, where the
OCR tool can be trained to recognize more languages or fonts, or to use the
knowledge of which character combinations are more likely to appear.
The better the image is, the more accurate the result gets. If the OCR
algorithm is applied on a horizontal text with very little or no noise and
    the
letters are well separated from each other it is possible to get a very good
result. But if the image is blurry, warped or noisy, or if the characters
    are
merged, it is likely that the output gets worse than it would otherwise. [1,
     2]
The output accuracy will also depend on the image resolution.
OCR software is generally better at recognizing machine writing than
handwriting. In machine writing the same character always looks the same,
given that the same font is used. With handwriting the characters differ a
    lot
depending on who is writing, and even if the same person has written a text,
the same character will be a little different each time it is written.
    Machine
writing often consist of more distinct characters, and since handwriting is
more varied than machine writing, it is harder for the algorithm to match
the characters to its database.
3
```

Output file 126: *OCR API Service – Image* n