# TEXT CROPPING SYSTEM

Group #5 Sponsored by: Dr. Amy Larner Giroux

# Group Members



# Project Goals

• Digitize Samuel Johnson's "A Dictionary of the English Language"

- 1755 and 1773 versions
- One of the most influential dictionaries
- Correct faulty scans and text warping
- Implement computer vision
  - OpenCV
- Solution could be used for future document preservation
  - Government archives, Museums, Universities

fand houfes, and allowing one man to every houfe, who could 2. To fhorten, to cut fhort. have any fhare in the government, (the reft, confifting of women, children and fervants) and making other obvious abatements, these tyrants, if they had been careful to adhere together, might have been a majority even of the people Swift on the contests in Athens and Rome. ollective 4. The caufe of abating ; extenuation.

As our advantages towards practifing and promoting piety and virtue were greater than those of other men; fo will our excufe be lefs, if we neglect to make use of them. We cannot plead in *abatement* of our guilt, that we were ignorant of our duty, under the prepoficition of ill habits, and the bials of a wrong education. Atterbury's Sermons. ABATEMENT, in law.

The act of the abator ; as, the abatement of the heir into the land before he hath agreed with the lord. The affection or paffion of the thing abated ; as, abatement of the writ. Covel. ABATEMENT, [with heralds] is an accidental mark, which being added to a coat of arms, the dignity of it is abafed, by reafon of fome flain or diffuonourable quality of the bearer. Diet. ABA'TER. n. f. The agent or caufe by which an abatement is

procured. *Abaters* of acrimony or fharpnefs: expreffed oils of ripe vegetables, and all preparations of fuch; as of almonds, pif-Arbithmot on diet. tachoes, and other nuts. ABA'TOR. m. f. [a law-term.] One who intrudes into houfes or land, that is void by the death of the former pofferfour, as we add, use is vow by the sears or cele former pomendi, is yet not entered upon or taken up by his heir. Dis. Austrupe. n.f. [ol records] Any thing diminished. Baily, A BATURE. n.f. [a having term.] Thole forgies of grafs which are thrown down by a flag in his pating by: Dis. Aus. n.f. The yarm on a waver's warp; a term among clo-

Chambers. A BBA.m.f. [Heb. 2N] A Syriac word, which fignifies father. A'BBACK. m.f. [Lat. ablatia.] The rights or privileges of an abbot. See ABBEY.

According to Felinus, an abbacy is the dignity itfelf, fince an abbot is a term or word of dignity, and not of office; and, therefore, even a fecular perfon, who has the care of fouls, is fometimes, in the canon law, alfo filled an abbot.

Ayliff's Parergon Juris Canonici. A'BBESS. n. f. [Lat. abbatiffa, from whence the Saxon abu-

byrre, then probably ablatefs, and by contraction ableffe in Fr. and abbefs, Eng.] The superiour or governels of a nun-Fr. and aboets, Lag, 1 nery or monaftery of women. They fled

Into this abbey, whither we purfued them ; And here the abbefs fhuts the gate on us,

And will not fuffer us to fetch him out.

I have a fifter, abbefs in Terceras, Who loft her lover on her bridal-day. Dryd. D. Sebaft. Conftantia's heart was fo elevated with the difcourfe of Father Francis, that the very next day fhe entered upon her vow. As foon as the folemnities of her reception were over, we re-tired, as it is ufual, with the *abbefs* into her own apartment. Addifon Speet. Nº 164.

ABBEY, or ABBY. n. f. [Lat. abbatia; from whence probably firft ABBACY; which fee.] A monaftery of religious per-fons, whether men or women; diffinguithed from religious houfes of other denominations by larger privileges. See ABBOT.

With eafy roads he came to Leicefter ; Lodg'd in the abbey, where the reverend abbot,

With all his convent, honourably receiv'd him. Shakefp. Henry VIII.

A'BBEY-LUBBER. n. f. [See LUBBER.] A flothful loiterer in a religious houfe, under pretence of retirement and aufterity.

This is no Father Dominic, no huge overgrown abbey-lubber; this is but a diminutive fucking friar. Dryd. Sp. Fr. A'BBOT. n. f. [in the lower Latin abbas, from ] # father; which fenfe was implied; fo that the abbots were called pawhen terms was implied to that the aboots were called pa-tres, and abbelies matter monoferit. Thus Fortunauts to the abbot Paternus : Naminit efficient jures, Paterne, geris.] The chief of a convent, or fellowthip of canons. Of thefe, fome in England were mitted, fome not : thole that were mitred, were exempted from the jurifdiction of the diocefan, having in themfelves epicopal authority within their precincls, and being alfo lords of parliament. The other fort were fuband being allo lords of parlament. The order for were full-ject to the diocefan in all fpiritual government. Coud. ABDU'CTION. n. [addudits, Lat.] I. The act of drawing apart, or withdrawing one part from

ABBY. See ABBEY.

A BBY. See ABBRY. A BBY, See ABBRY, A C A particular form of argument. To ABBRFV IATE. s.a. [Lat. aldereinra] To Thome by contraction of parts without lofs of the mini-the function of the second secon To ABBREVIATE. v. a. [Lat. abbreviare.] I. To thorten by contraction of parts without loss of the main

It is one thing to ablreviate by contracting, another by cut-Bacon. Effay 26. ting off. The only invention of late years, which hath any way contributed towards politenets in difcourfe, is that of abbreviating or reducing words of many fyllables into one, by lopping off Swift's Introduction to genteel conversation. the reft.

Against this opinion we may very well fet the length of their days before the flood; which were abbreviated after, and in half this fpace contracted into hundreds and threefcores. Brown's Vulgar Errours, b. vi. c. 6.

ABD

ABBREVIATION. n. f. 1. The act of abbreviating 2. The means used to abbreviate, as characters fignifying whole

words Such is the propriety and energy of expression in them all, that they never can be changed, but to diadvantage, ex-cept in the circumftance of using abbreviations.

Swift's Introduction to genteel conversation. ABBREVIA'TOR. n. f. [abbreviateur, Fr.] One who abbreviates, or abridges.

ABBRE'VIATURE. n. f. [abbreviatura, Lat.] 1. A mark used for the fake of fhortening.

2. A compendium or abridgement. He is a good man, who grieves rather for him that injures him, than for his own fuffering; who prays for him, that wrongs him, forgiving all his faults; who fooner fhews mercy than anger; who offers violence to his appetite, in all things endeavoiring to fubdue the flefh to the fpirit. This is an ex-cellent abireviature of the whole duty of a chriftian.

Taylor's Guide to devotion ABBREUVOIR. [in French, a watering-place. This word is derived by Menage, not much acquainted with the Teutonic dialects, from adbibare for adbibare; but more probably it comes from the fame root with brew. See BREW.] It fignifies, among mafons, the joint or juncture of two flones, or the interflice between two flones to be filled up with mortar. Dieta A, B, C.

 Is taken for the alphabet; as, he has not learned his a, b, c.
 Sometimes for the little book by which the elements of reading are taugh

To A'BDICATE. v. a. [Lat. abdico.] To give up right ; to refign ; to lay down an office. Old Saturn, here, with upcaft eyes,

Addilon Beheld his abdicated fkies. ABDICA'TION. n. f. [abdicatio, Lat.] The act of abdicating; relignation ; quitting an office by one's own proper act before

the ufual or flated expiration. Neither doth it appear how a prince's abdication can make any other fort of vacancy in the throne, than would be caufed by his death ; fince he cannot abdicate for his children, otherwife than by his own confent in form to a bill from the two houfes. Swift on the Sentiments of a Church of England man. A'EDICATIVE. adf. That which caufes or implies an abdica-Dia

Shakefp. Comedy of Errours. A'BDITIVE. adj. [from abdo, to hide.] That which has the power or quality of hiding. Dist. ABDO'MEN. n. f. [Lat. from abde, to hide.] A cavity com-

monly called the lower venter or belly : It contains the fto mach, guts, liver, fpleen, bladder, and is within lined with a membrane called the peritonæum. The lower part is called the hypogaftrium ; the foremost part is divided into the epigastrium, the right and left hypochondria, and the navel; 'tis bounded above by the cartilago enfiformis and the diaphragm, fideways by the fhort or lower ribs, and behind by the vertebræ of the loins, the bones of the coxendix, that of the pubes and os facrum. It is covered with feveral mufcles, from whole alternate relaxations and contractions in refpiration, digeflion is forwarded, and the due motion of all the parts therein contained promoted, both for fecretion and ex-Quincy. The abdomen confifts moreover of parts containing and Wifeman's Surgery.

contained. ABDO'MINAL. & adj. relating to the abdomen.

To ABDU'CE. v.a. [Lat. abduce.] To draw to a different part ; to withdraw one part from another. A word chiefly ufed in phyfic or fcience.

And if we abduce the eye unto either corner, the object will not duplicate ; for, in that polition, the axis of the cones remain in the fame plain, as is demonstrated in the optics de-Ilvered by Galen. Brown's Vulgar Errours, b. iii. c. 20: ABDU'CENT. adj. Mufcles abducent, are thole which ferve to open or pull back divers parts of the body ; their oppofites be-defined abducent of the body ; their oppofites be-put data and the second secon ing called adducent.

another.

members. In purfuance of this theory, he fuppofed the confirictors of the cycleids muft be firengthened in the fupercilious; the *ab-ductors* in drunkards, and contemplative men, who have the fame fleady and grave motion of the cye. Arbuthnot and Pope's Martinus Scriblerus.

#### Average Page

- Guide word at top of column
- Definitions spanning each column

ABORI'GINES. n. f. Lat. The earlieft inhabitants of a country; those of whom no original is to be traced; as, the Welsh in Britain.

To ABO'RT. v. n. [aborto, Lat.] To bring forth before the time; to mifcarry. Dist.

ABO'RTION. n. J. [abortio, Lat.]

- 1. The act of bringing forth untimely.
- 2. The produce of an untimely birth.

His wife mifcarried; but as the *abortion* proved only a female foetus, he comforted himfelf, that, had it arrived to perfection, it would not have anfwered his account.

Arbuthnot and Pope's Martinus Scriblerus. Behold my arm thus blafted, dry and wither'd, Shrunk like a foul abortion, and decay'd, Like fome untimely product of the feafons, Robb'd of its properties of ftrength and office. Rowe's Jane Shore.

# Parts of a definition

- Each orange box is an entry
- Yellow box is a headword (beings every entry)



### Extraneous Information

 These parts of the page are to be excluded from the definition

# Project Requirements

- Analyze each full page image for page level rotation and correct for it
   Parse each page into word-level image files based on the headwords
   Analyze the word-level images for rotation and correct for it
- Combine multi-part word images into single files of common width to keep text size common between the parts
- Adjust contrast and brightness of multi-part words to make the image more uniform
- Name the resultant file based on our standard naming conventions and save in a directory structure arranged by edition (1755/1773) and beginning letter of the word

# Design Approach

- Load image(s) into the program
- Split pages into left and right columns
  - Allows for easier image dewarping
- Dewarp image
- Identify headwords at the beginning of each entry
  - Each entry begins with a capital letter and is followed by definition
- Split each word into its own file
  - If definition spans column or page stitch together
- Save



Project Block Diagram



Program Execution Flowchart

# GUI

#### • 4 Functions

- Choosing a directory
- Choosing an image
- Saving to a location
- Process the images

Text Cro	opping System v2		-	□ ×
	Iext (	Cropping	System	
	Directory Choose	I	Select Image Choose	2
		Save		
		Process		

# GUI

In this window, the user is able to select a directory of images to process.



# GUI

In this window, the user is able to select a specific image.

The file filter could also be adjusted.



# Output and PyTesseract OCR

Desired output is as follows:

- Each definition should be cropped to include every definition for a headword.
- Each entry must be saved into a .tif file under the following naming system:
  - So a complete file name would be "f1755-abacus.tif".
  - Multiple definitions would be appended with a "-#", # = occurrence
- PyTesseract will be used to read the headword and name the files as such.

# File Naming Successes

ABATEMENT, in law. The act of the abator; as, the *abatement* of the heir into the land before he hath agreed with the lord. The affection or paffion of the thing abated; as, *abatement* of the writ. Cowel.



ABATEMENT, [with heralds] is an accidental mark, which being added to a coat of arms, the dignity of it is abased, by reason of some stain or dishonourable quality of the bearer. Dict.

f1755-ABATEMENT-0

# Image Dewarping

- Two processes of the dewarping algorithm
  - Text span detection
  - Image remapping





Then were the first further and a pipeters, command-ing all fach a had book in the book in the same is give the co-upon. Sir John Davies on tretery.

the most status power of prince connot alfabera. Laste 3. 10 pronunce a fin remark d, in the end of the d for fe-

To promine a fin manual in the exploring in the field of the field of

becaufe er thing, the proceed from his Decause and an of the proceed from him one point which the which is a second with the second second second second second field have for particular of that they are sum to be finded have for particular of that they are sum to be tend. Association of the second second second second second Although it rung in format second second second second second it is pronounced. Second second second second second second 3. Not reflecting an area of the second second

4. Not limited ; as, abfalate power. My crot m is abfalate, and holds of none ;

Without relation. Algebracky we cannot different and its entropy algebracky ap-performance of the second standing trading Hester, b. v.

Partic than being the pathenal carfor of real, the greatest peak or the greatest will, either adjuinely for in the station or attively for the real is in the scalar of the scalarity affected to the one against the other. Sprai is Sermen. 1 Bit fails with the state of the constraints of the con

ABS

citler to be done or avoidel. Reicht", Bregad.
 c. Persporting, politively.
 compositive politively.
 compositive politic for not togat
 Compositive to the state of the state of

To findly, but the subject of the subj

 The only data of the state of t from the purpofe.

Information of a sublidity on the first state of a sublidity of a sublid

My crown is addedute, and held of more provide the source of a scalable constraint, the flow of the source of the

A south a set of the set A Moticine that, by the former or poonty or the parag-cilder cash, the specified or paragent because, a dick a way fupperflower motifure in the body. Quinty, There is a third diffs of fullance, contrology called any there is a the grines thick of fulls, could, called any there is a the grines thick of fulls, could, call a sub-enty is at the grines the or effectivenes, and we herefore called

### Image Dewarping cont.

- Identify contours
- Connects the contours into spans

in party forest their wass na where the first fickets we where in ucose of the weath.

Sights Milestan Sights Milestan at a provideo ndo that blocky with la ball oth. Mallo Lingid's fr

Alt, the gund av ben Mina's S. Ana/e Ly/t, b. Vis & 90 Spiriter, Lat ]

tere 2 2.3 the regard out are being has to chart . High's Legic. In ais lenfe ve loak of the ableive can abfelve in grammar.

In an endowe first of the abbrive cap abbrar in grunnas, one lipitions are capable of the abbrive and abbrar in grunnas. Abbrive cosen is abbrive and obtain a second seco

Which then he some if the finate here in file words, "Anal built of fending space while the fifth of "Anaw words," and the space of the space When a space of the the space of the the space of the spac

the adjate/y dime they ark of virae, nor finally fink time the first events. Sidney When main thready haild upon having the covering a pre-tifiant any source is an end of the source of the source with and the source of the source of the source of the source adjates are source only back for each of the source adjates to the source of th

..... 8 8 . . . . .

ang time had countral formuse's

And of that paper, for the floft poto se things abfected into all rights favorion necessary, either to be hed or desice

Books's gregas. renptorily, positively. I being a Lad, why dea not then commund any effective not to sa Going into fuel dange, a you hight? Poad. Laft, b. is

BSOLD PENLISS. n. f. [fros abfelge.]

of and in the comprision was ge for mich an deen upon ober. Suit's See

Convertier, and the second sec

a noto vroze mje popularstvo popular na objekt sequites of a fail acquites of a fail acquites of a fail acquites of the sequence of the seq May, when ad among divino it likewise figning a release on of his got the oblighter of the spin terms pronound ther in a case of law, or got in for spin third. The three is, in the lang of law, or got kindle dividuality the production of the spin terms of the spin terms of the billicity, and another filled a declaptory or excepting of the spin terms of the spin terms of the spin terms of the lifeting of figs, or may not desired by cells this atthem.

Die advision promousers by a prieft, so which popilit or morting and a contrast of the ground to save the po-trast of the second second second second second the second second second second second second second an advised second seco

A main that will be added a more than a more of the second and the s

ABS

1 Que the first figure And a Quest abjaces, continiand Sir John Device on Irelan h s i ov ov the profits (i) out of Leen hing. from able the profits (i) out of Leen 9518, Su (PSIS) Austration (alights, Lat.) Yo fand off, in laye of 207. A 1850 V V. v. a. [alights, Lat.] To char, in acquines a cristiant zjudicial faithe

You you good the strate of the is a maximum structure of references in registrate grant structure of references in registration of the structure of the structure in the structure is blocky own, the structure is blocky own, in splok by by bah. Parlow is Mail & Prog., which and needed by comparison the per-ture of the structure of the structure of the structure is a structure of the structure of the structure is a structure of the structure of the structure is a structure of the structure of the structure is a structure of the structure of the structure is a structure of the structure of the structure of the structure is a structure of the structure of the structure of the structure is a structure of the structu

dience to the sociffate, and from whi

an, abjig og Biltes heg. igh is fo fuppoled infinitely differt from whe is not diffant from us by a unite in a val, and not infinite-t one circulation which precoded it, and out ne-

ours, and conference by abfolved in the fpage of What Bufe

entity, fo late to build u the work botton, here foon Milite's Passafe Left, b. vii. 1.94 [abfellur, Bat.] bild as Well to perform as mind.

abfoldur, [at.] by chell a perform as hingl. the proceed from his we perform, with-f defect or main; it equate the but the the

6. Unonditional ; es, an abfelete premine South's Serions.

Not relative; as, adjudanc price; I are full the dilinguinon of four-ray and influence, as a price and relative sporthing, will be ray have out in the writing of any source of the relative to call, as did at leafly it idont in any source of the relative to call, as did at leafly it idont in the working of share. Stilling all of of of the Res. Idol. Any adjust bodies in the sufficiency of the leafly of the edge of the regard that the beings of the being the route of the adjust of the regard that the beings of the being the route of the adjust of the regard that the being the route of the route of the being the

In this first the fight of the table we get a state of the state of th

Long is it fines I faw binn, By time heft nothing blogd that line generation of the second and burt of pecking were a his 1 m adjágar.
Was a histgenet by land 2.
Grand and merging: by the second merging by the secon

Compressive withour reflection.
 All the characterization of the second sec

Without relation. Abjettely or come diformenend, se annot abjeturgs ap-appendix of the second sec

ABS

ithout lists or dependence. The prince long time had childed fortune's live aut, once points d, did *abfolgely* reign: those with their Ahnzons, the herein firove, and conque'd firft thole beaution gos world some

Withou confliction. And of this narrow, for the and rates are beings of fairly into all more fairgeon necessary, either to be held or denied, either to be done or avoided. S. Plermetorily, politively. Henge 1.5. why wild not then formand are afforded not then the formation into danger. when fairl? P and Lage, b. in A source exact, m. f. [Iron afforder.]

orbitance, indoff indo derendance, or lifete. The allocitant and ullighted is of his contention was ge-glib pluep hoteen of. They is nothing that can Alife a Agen to that guerous ad of condition, so each the to ringe, to faun, or to do only; but that which give him that hopping's with-elf, for which are depend upon once. South's Sec.

espoticife. We kept a figuit hand on his nobility, and chole fithe fi Wance choice a part and on his hobins, and choice with a Viance choice may and havers, which we can collections to hid, but had blas intereft in the peede: which cade for his abjutacif, but not for his frice. BOOLD NON. n.f. (Affalatis, Lat.)

Otherity. The adjoints pronouncil by a prior, thehn point or order that a series in infailled ground to save the pri-ton, fo ablowd, on for the convert group. Save a save or an adjoint of the save of the save of the save the save of the periors, upon the account of pronounce in yet, if adjust that around the prover, he may be again proceeded near the other of the save of the save from the purpose. The save of the save of the save of the save of the periors, the save of the save of the save of the save of the periors. The save of the save of the save of the save of the periors, the save of the s

2. To luck up. See Ansonment. Suppose the forementioned confusion final density of the grabit as to adjust and attenual the faid foregoing page to an entropy degree, it is crident, that the found sental part and negativity come into danger. Locar as Conformation. A sufficient that, but to forefore on configs of its pare, either cause the perities of pungent bebudg, offices away fuperfluous moliture in the body. Quinty, There is a third (do of fublicnet, connends) and an entropy of the visuous kinds of fuels, coral, challs, office or parts; as a the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challs, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of fuels, coral, challe, office or entropy of the visuous kinds of the

bents; as, the various kinds of facilis, coral, challs, craps even

### Image Dewarping cont.

Identify keypoints • of text and apply coordinate remap

Aquint: Applied, in the soil by, thooks full acquinted if profon by tome and former of law; also, a temporary dis-curse of he are a stempting or any in provide through the or these and according to the origination of the there or these in pleasings; it does its release or of his grade the colling of fore former product the rest of the full of the origination of the former of his field of the origination of the former product there is, in the full of the origination of the there is, in the full of the origination of the plate the full of the origination of the former of the full of the full of the origination of the plate the full of the origination of the former of the full of the full of the origination of the former of the full of the origination of the full of the plate of the full of the origination of the former of the full of the origination of the full of the origination origination of the origination of the origination or the origination or the origination or the origination origination or the origination origination or the origination origination or the origination origination or the originatis origination ore origination origination origina

A'BSONOUS, adj. [abfonus, Lat. ill-founding.] Abfu

To happed an uniter of a middle confliction, the final article of fonts of the qualifies of both, unwarranged by any of a facting, y. Solt allower to our maion. Colorall's Second Secon

Dryden. 1. An humming through their waxen city grows. 2. To make an inarticulate and buzzing found. I think he'll hear me : yet to bite his lip, And hum at good Cominius, much unhearts me. Shakefp. Upon my honour, fir, I heard a bumming, And that a firange one too, which did awake me. Shakefp. The cloudy meffenger turns me his back, And bums ; as who fhould fay, you'll rue. Shakefp. Macbeth. 2. 3. To paufe in fpeaking, and fupply the interval with an audible emiffion of breath. Having pump'd up all his wit, And humm'd upon it, thus he writ. Hudilras, p. iii. I still acquieft, And never humm'd and haw'd fedition, Hudibras, p. iii. cant. 2. Nor fnuffled treafon. The man lay humming and hawing a good while; but, in the end, he gave up himfelf to the phyficians. L'Eftrange. Still humming on, their drowfy courfe they keep, And lafh'd fo long, like tops, are lafh'd afleep. Pope. 4. To fing low. The mulical accents of the Indians, to us, are but inarticulate bummings; as are ours to their otherwife tuned or-Glano. Apol. gans. Pope. Hum half a tune. 5. To applaud. Approbation was commonly expressed in publick affemblies by a hum, about a century ago. HUM. n. f. [from the verb.] 1. The noife of bees or intects. To black Hecat's fummons The fhard-born beetle, with his drowfy bumit, Shake (peare's Macheth. Hath rung night's yawning peal. Nor undelightful is the ceafelefs hum, To him who mufes through the woods at noon. Themfon. 2. The noife of buffling crowds. From camp to camp, through the foul womb of night, Shake/p. Hen. V. The hum of either army ftilly founds. Tower'd cities pleafe us then, Milton. And the bufy burn of men. One theatre there is of vaft refort, Which whilome of requefts was call'd the court ; But now the great exchange of news 'tis hight, And full of hum and buz from noon 'till night. Dryden. 3. Any low dull noife. Who fat the neareft, by the words o'ercome, Slept faft; the diftant nodded to the hum. Pope's Dunciad.

but if night-robbers lift the well-itor d hive, An humming through their waxen city grows. Dryden. 2. To make an inarticulate and buzzing found. I think he'll hear me : yet to bite his lip, And hum at good Cominius, much unhearts me. Shakefp. Upon my honour, fir, I heard a bumming, And that a firange one too, which did awake me. Shake/p. The cloudy meffenger turns me his back, And hums ; as who fhould fay, you'll rue. Shakefp. Macheth. 3. To paufe in fpeaking, and fupply the interval with an audible emiffion of breath. Having pump'd up all his wit, And humm'd upon it, thus he writ. Hudilras, p. iii. I ftill acquieft, And never humm'd and haw'd fedition, Nor fnuffled treafon. Hudibras, p. iii. cant. 2. The man lay humming and hawing a good while; but, in the end, he gave up himfelf to the phyficians. L'Eftrange. Still humming on, their drowfy courfe they keep, And lafh'd fo long, like tops, are lafh'd afleep. Pose. 4. To fing low. The mulical accents of the Indians, to us, are but inarticulate hummings; as are ours to their otherwife tuned organs. Glano. Apol. Hum half a tune. Pope. 5. To applaud. Approbation was commonly expressed in publick affemblies by a hum, about a century ago. HUM. n. f. [from the verb.] 1. The noife of bees or infects. To black Hecat's fummons The fhard-born beetle, with his drowly burnt, Hath rung night's yawning peal. Shake peare's Macheth. Nor undelightful is the ceafelefs hum, To him who mufes through the woods at noon. Thamfan. 2. The noife of buffling crowds From camp to camp, through the foul womb of night, The bum of either army ftilly founds. Shake/p. Hen. V. Tower'd cities pleafe us then, And the bufy burn of men. Milton. One theatre there is of vaft refort, Which whilome of requefts was call'd the court ; But now the great exchange of news 'tis hight, And full of burn and buz from noon 'till night. Dryden. 3. Any low dull noife. Who fat the neareft, by the words o'ercome, Slept faft; the diftant nodded to the hum. Pape's Dunciad.

### Image Dewarping cont.

- Left is without rotation before dewarping
- Right has rotation applied before

# Successes and Difficulties

- Successes:
  - Open source software solution
    - Worked well without modification
    - Easy to modify
- Difficulties:
  - Getting started
    - Minimal knowledge
    - Many programs found did not work
  - Debugging

# Column Cropping

- Before cropping the dictionary entries, the columns must be cropped to reduce errors.
- Any excess whitespace and unwanted text (page numbers, guide words) is removed in this stage.
- Cropping the columns also provides us with the width of the definitions.
- The page on the right will be used as an example throughout this section.

ABO ABL There is a natural analogy between the addative of the hody and the purification of the foul, between eating the hody bread and drinking the forsel challers and a participation of the body and blood of Chrift. *Taylor III arthy Communicano*, Walki by the being wares, the pious train Are cleansid, and calt the aldotions in themin. *Poor I lind*. The ranking of dominal preparations in water, to dilibite and suff take the teft ; which is an abjuration of fome doctrines of church of Rome. . There is likewife another oath of abjuration, which laymen clergymen are both obliged to take ; and that is, to abjure retender. Ayliffe's Parergan Juris Canamici, tLACTATE. v. a. [ablaite, Lat.] To wean from the wash away any acrimonious particles.
3. The cup given, without confectation, to the laity in the popilities. n. f. One of the methods of grafting; and, the fightfication of the inclusion of grading, and, the fightfication of the word, as it were a wear-n by degrees from its mother flock, not cutting from the flock, till it is firmly united to that on churches. To A'INPECATE e. s. [from sharps, Lat.] To deny, Arstee Artose, n. / [sharps, Lat. denial, from sharps, to The sharps and the start of the sharps of the sharps of the tereflex, and raths of all dura man is anoth pay to depend your, dath le may the more expedibly how Church. Arston Artons, n. / [sharps, Lat.] The ski of catting any how from tere is a term of graneleng. Dir. hich it is grafted. AQUEA'TION. [ablaquatis, Lat.] The act or practice of ening the ground about the roots of trees, to let the air and operate upon them. ench the ground, and make it ready for the fpring : Pr alfo foil, and ufe it where you have occafion : Dig bo Reserve as yet note of they, where attemption is regardly a set of the set Uncover as yet roots of trees, where ablaque ABNODA'TION. m. f. [annuality and f. Ditt. knots from trees; a term of gardening. Ditt. ABNO'RMOUS. adj. [abnermin, Lat. out of rule.] Irregular, Ditt. mithagen. Did. Richa R. D. and (a. [a fasterm, but adopted into common language, derived immediately from the French a length, and die a length, and segme a length. Breaks infild a word of very dochted a elingth, and engen a length. Breaks in the an orient bases, fignified a length, in which fends, seg a length, is to the up orthother in a thin.] In a flip. Which, when far off, Cymocles heard and faw, He loadly call? It for that were abardy. The Body call? It for the wore abardy. And him to there over that along the Fare of the Linear 6. In the line bark who no in horse to draw, And him to freely over that deep ford. Fairy  $Q_2$  b, ii. cant. 6. I made this answer, that he might land them, if it pleafed him, or otherwisk keep them absord. Sin *W*. Reaulight Elegrit, When morning role, I fent my mates to bring Supplies of water from a neighbring fring; Whill I the motions of the wind, explored a ' Then fummon'd in my crew, and went absard. Addifon's Ovid's Metamorphofes, b. iii. And by: a. f. (from whid.) And by: a. f. (from whid.) . Habitation, dwellings place of refidence. . Bait I know thy and and thy going out, and thy coming in, But I know this may a strong star are a strong s But I know thy about and thy going out, and and thy rage againft me. 2 Others may use the ocean as their road, Only the English make it their about ; Whole ready fails with very wind can fly, And make a cov nant with th' inconftant fky. d an *able* body, for the which the prince admits him : for the ince himfelf is fuch another : the weight of an hair will turn s bakefp. Harry IV. p. ii. 2. Stay, continuance in a place. 1. Suby, continuance in a pace. Sweet Finds, your patience for my long abide; Not I, but my affairs, have made you wait. Subsifyener I: Marchant of Venier, Making a floort abide in Sicily the fecond time, landing in Italy, and making the war, may be reasonably uside the budi-nets but of ten months. Dryder's Dedicat. to Encid. The workersky subsidier and abid. There have been fome inventions alfo, which have been The woodcocks early vifit, and abede The woodcods early vifi, and add Of long contrast, in our temporate classe, Partle 3 liberal harved. Partle 3 liberal harved Theory contrast of the second second second Theorem and the second second second second second second the second second second second second second second second the second second second second second second second second the second second second second second second second second the second second second second second second second second the second second second second second second second second the second second second second second second second second the second second second second second second second second second the second le far the utterance of articulate founds, as the fpeaking of rtain words. Wilkins's Mathematical Maric. early for the utreame of articular founds, and the facking of excitation work. The second second second second second second commonly ufield. See Exants. Part for with gold, And the fluxing lines of julice hardful breaks: New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second second second second second New does of the second the second sec The fullar force of a Starfport Hery VII. Ref Orsasters. A for from stable JA forcet anticipation of founding future in impedient upon the mind of form event to come i programmed the star of the star-ter may men that fumble at the therhold, Are will formed that diagraph the within — — Tubi I may, additionari mult not now affinget us. My fock lifting, being furnershare include, star but for form to affin have the star of the star of the star of the star My fock lifting the the data is an affinist on a star of the star of the star of the star of the star to affinish the data is an affinish form advance may will it Nov equation that does not affinish form advance may will it is a star of the star of the star of the star To ABOUTISH, w.e. from using Lat. to blot ouz] Shakefpeare's Henry VU! that one is weary on. BLEGATION. n. f. [from ablegate.] A fending abroad, or out Did. As a GAYLON, K. J. [From any first processing and the second processing of the second processing An action is the original and the approximation of 10 not provide the method of the second seco 
 Diff.
 establish
 Charmann, k.viii.

 sy.1
 r.
 To pot an end to; to defrey.
 Charmann, k.viii.

 syr.1
 The long continued wars, between the English and the Scots, between the bar folds the fract scots of the scots of the scots.
 Diff.

 Diff.
 balt then raifed invisible schoolings and hats, which long continued percee hath fince abs/field.
 Sur folds the fractional scots of the sco That which wafnes clean, That which has the power of cleanfing,

#### Otsu's Thresholding and Binarization $\rightarrow$ Dilation to find contours $\rightarrow$ Accept contours if they are large enough

#### ABL

ATE, v. e. foblatte, Lat. | To wea

TO A'BLEGATE.

ritis, Lat.] A prodical i

АВО

thy coming in Kinet, xix, 27



ift take the teft ; which is an abjuration of fome doctrines of he church of Rome. There is likewife another oath of abjaration, which laymen and clergymen are both obliged to take ; and that is, to abjure he Pretervide: ABLACTATE v. a. [ablacts, Lat.] To wean from the breall: BACTA'TION. n. f. One of the methods of grafting; and, according to the lignification of the word, as it were a wean-ing of a cyon by degrees from its mother flock, not cutting it of wohly from the flock, till it is firmly united to that on it on woony room the nock, in first mindy antice of and on which it is grafted. BEAQUEATION. [ablaguestis, Lat.] The set or practice of opening the ground about the roots of trees, to let the air and water operate upon them. opening the ground about the root of irrers, to let the size and warr openize apout them. The size results of the first 2. The size of the first 2. The size of the first 2. The size of the first 2 models of the size root is of the size with the size of the size nly this yearly fruit will much decreafe from time also the whole body and boughs of that precious hadrogram and only in yorky fractional mode detection from time two its idity will also hadrogram of the star of the star two its idity, will also hadrogram of the star of the star two its idity, will also hadrogram of the star of th and the strength of the source of the barrier of an line of the source of As LEO ACTION:  $m_{eff}$  (from able) to ensuing measure to Dill. O the way. That mation do the locked, both for correlations and alwayf, that from neighbour contrain they enfantify come, fonce to firme, fonce to learn, fonce to behold. Subary, k in America and analytication. Carl American Dill, medy allo unadividual. BLIGURI'TION, n. f. [abliguritis, Lat.] A prodigal fpending on meat and drink. Did. on meat and drink. Fo A'BLICATE. v. a. [akligs, Lat.] To bind or tye up from. D. To A'BLOCATE. v. a. [akligs, Lat.] To let out to hire. Perhaps properly by him who has hired it from another. Cabini Laring Turing Antoc A'TIO', m.f. [from allowic, A letting out to hire, To Antu/un, w.m. [allow], Lat.] To be unlike. Di. A'turuser, and [allow], Lat.] To be unlike. Di. J. That which has the power of cleaning. Di. Antu/vinov. m.j. [allowing, Lat.] I. The allo diaming, or wallang clean. Dia.

There is a narpid galaxy between the addocing of the hadroging the population of the forst phenomenon entry the hadroging and the present of the forst phenomenon of the body phenomenon (the forst phenomenon) and the phenomenon (the pheno 3. charding layor introductionary does any in an optimal To AUNECAT E. o. a. (from along), La. I. To dery, Dawn Cortex, w. f. (alonguint), Lat. denials, from along, to dery.] Deail, remainstant, and the own bolks and along the strength of all Darman's is and get to dery and using the strength of all Darma's is and get to dery and using the strength of the Hamman J. Pathibial Catalyja, Assoc o'rrow, w. f. (alonguint), and J. To along the strength loss from trees i a strength of gethering. Assoc or trees, where the strength of the strength of the strength loss from trees i a strength of gethering. A matching and the probability of the probability Then fammed in my creen, and were a lower density of the state of Malling Color Matemptifying, is it have the state of Malling Color Matemptifying, is have the state of the state of the state of the state and the grage significance of the state (b) and the state of the state bound of the state of the state of the state of the state bound of the state of the state of the state of the state bound of the state of the state of the state of the state bound of the state of the state of the state of the state bound of the state of the state of the state of the state bound of the state of the state of the state of the state bound of the state of the s and but of consistence . Depicts' Distance set for Of long contrasts in our template set of the South South

to alk num, whether ne nan never any lecter abasement mind? No, replied the duke; but I think fome adventue kill me as well as another man. To ABO'LISH. v. a. [from absles, Lat. to blot out.]

De ABO Dient: w.e. (non-nearly the stabilished, were prefamilied). For us to aksign what he hash effahlished, were prefamilied in the finite of the most index to the parliament's part if was proposed, that all the hitforge-dense, and chapters, might be immediately taken away, and absigned. Currents, Aviii. densi, and chapters, might be immediately taken away, and abilitied. To top at an end to; to deffroy. The long continued wars, between the Englith and the Scots, had then raifed invincible jealoins and hats, which long con-tinued peace hath funce abilitied. Sir Jein Hoymand.

# Entry Cropping

How can the entries be isolated? Unique features.

- Unindented text is either a headword or a number denoting a new definition.
- Headwords are always capital letters only, except for when it starts with "To" (with other minor exceptions).
- Headwords are sometimes written in a larger font.
- Entries always span from the top of one headword to the top of the next. (Obvious, but necessary for determining bounding boxes. The width of the definitions is already determined through the column cropping.)

To A'BNE'GATE. v. a. [from abnego, Lat.] To deny. ABNEGA'TION. n. f. [abnegatio, Lat. denial, from abnego, to deny.] Denial, renunciation.

The abnegation or renouncing of all his own holds and interefts, and trufts of all that man is most apt to depend upon, that he may the more expeditely follow Christ.

Hammond's Practical Catechifm.

ABNODA'TION. n. f. [abnodatio, Lat.] The act of cutting away knots from trees; a term of gardening. Diet. ABNO'RMOUS. adj. [abnormis, Lat. out of rule.] Irregular, mifhapen. Diet.

ABO'ARD. adv. [a fea-term, but adopted into common language; derived immediately from the French à bord, as, aller à bord, envoyer à bord. Bord is itfelf a word of very doubtful original, and perhaps, in its different acceptations, deducible from different roots. Bond, in the ancient Saxon, fignified a house; in which fense, to go aboard, is to take up refidence in a fhip.]

In a fhip.

Which, when far off, Cymocles heard and faw,

He loudly call'd to fuch as were aboard,

The little bark unto the fhore to draw,

And him to ferry over that deep ford. Fairy 2. b. ii. cant. 6. I made this answer, that he might land them, if it pleased him, or otherwise keep them aboard. Sir W. Rawleigh's Essays.

When morning rofe, I fent my mates to bring

Supplies of water from a neighb'ring fpring;

Whilft I the motions of the winds explor'd;

Then fummon'd in my crew, and went aboard.

Addison's Ovid's Metamorpholes. b. iii.

# Entry Cropping

The first prototype for cropping the entries takes the very left side of a column and compares the size of everything there to determine if it is:

- A capital letter (A headword)
- A number (Additional definitions for a previous headword)

Challenges with this include some numbers being too large and some headwords being too small. This can be caused by a number of things including:

- The page was not flat on the scanner during the scanning process.
- Thresholding either including additional specks (shown on the 1), or failing to properly threshold a letter due to brightness of the image.
- The dewarping process modifying the size of the text.



# Data on Entry Cropping

These results are from testing on five different (and fairly straight) columns.

Total Headwords	Correctly Determined	Correctly Determined	False Negative	False Positive
and Numbers	Headwords	Numbers	Headwords	Numbers
97	49	30	11	7

Total Correct	Total Incorrect	Success Rate
79	18	81.4%

A

2.

# PyTesseract Entry Cropping

Initial ideas for PyTesseract involved applying it to the thin columns of text to determine whether the character is a capital letter or a number. However, PyTesseract does not work well when given the thin column of text.

Our next idea was to feed PyTesseract the entire cropped column, and have it determine if the letter is far enough to the left, and if it is a capital letter. (Close to what was done before, but we let PyTesseract have the entire column.)

PyTesseract training can be done for the specific font used by the dictionary.

Results from testing on the same five columns as before:

Total Headwords	Total Correct	Total Incorrect	Success Rate
61	59	2	96.7%



# Entry Cropping Precision

- Knowing where the headwords are is only the first step to the problem, as cropping based on the headword will cut off any letters that are higher on the same line.
- Once a headword is found, the entire line must be checked to determine where the highest point is.
- Similarly, the end of the definition must be checked to determine where the lowest point in the definition is.

A'BBEY-LUBBER. n. f. [See LUBBER.] A flothful loiterer in a religious houfe, under pretence of retirement and aufterity. This is no Father Dominic, no huge overgrown abbeylubber; this is but a diminutive fucking friar. Dryd. Sp. Fr. A'BBEY-LUBBER. n. f. [See LUBBER.] A flothful loiterer in a religious houfe, under pretence of retirement and aufterity. This is no Father Dominic, no huge overgrown abbeylubber; this is but a diminutive fucking friar. Dryd. Sp. Fr.

# Image Stitching

•Not every entry in the book remains on a single page or column.

•To account for this, we have need of a stitching program to take two different pages and combine a cropped image from both

# Image Stitching

•This will be done by adding onto the first image file. As the entry is cropped it will save into this file when it reaches the end of the entry, page, or column. If it is the end of the page it will repeat this process for the next page or column.



# Brightness Manipulation

•Not all the pages retained the same background color through the years

•Some page backgrounds were darkened enough during scanning to hurt readability

The first letter of the European alphabets, has, in the English language, three different founds, which may be termed the broad, open, and flender. The broad found refembling that of the German a is found, in many of our monofyllables, as all, wall, malt, falt; in which a is pronounced as au in caufe, or aw in law. Many of these words were anciently written with au, as fault, waulk; which happens to be still retained in fault. This was probably the ancient found of the Saxons, fince it is almost uniformly preferved in the rustic pronunciation, and the Northern dialects, as maun for man, haund for hand. A open, not unlike the a of the Italians, is found in father, rather, and more obscurely in fancy, fast, &c.

# **Brightness Manipulation**

By adjusting the contrast on these pages we can achieve a more uniform background while making pages easier to read

> The first letter of the European alphabets, has, in the English language, three different founds, which may be termed the broad, open, and slender.

The broad found refembling that of the German a is found, in many of our monofyllables, as all, wall, malt, falt; in which a is pronounced as au in caufe, or aw in law. Many of these words were anciently written with au, as fault, waulk; which happens to be flill retained in fault. This was probably the ancient found of the Saxons, fince it is almost uniformly preferved in the rustic pronunciation, and the Northern dialects, as maun for man, haund for hand.
 A open, not unlike the a of the Italians, is found in father, rather, and more obscurely in fancy, fast, &c.

The first letter of the European alphabets, has, in the English language, three different founds, which may be termed the broad, open, and slender.

The broad found refembling that of the German a is found, in many of our monofyllables, as all, wall, malt, falt; in which a is pronounced as au in caufe, or aw in law. Many of these words were anciently written with au, as fault, waulk; which happens to be still retained in fault. This was probably the ancient found of the Saxons, fince it is almost uniformly preferved in the rustic pronunciation, and the Northern dialects, as maun for man, haund for hand.
A open, not unlike the a of the Italians, is found in father, rather, and more obscurely in fancy, fast, &c.

# Unsharp masking

- To improve the readability of pages with blurring we use a technique called unsharp masking
- This creates a blurred copy of the image and applies it as a negative to sharpen the image



# Complete image enhancement

•Several pages have discolorations on them

•our program reduces this along with other noise elements to improve readability

The first letter of the European alphabets, has, in the English language, three different founds, which may be termed the broad, open, and slender.

The broad found refembling that of the German a is found, in many of our monofyllables, as all, wall, malt, falt; in which a is pronounced as au in caufe, or aw in law. Many of these words were anciently written with au, as fault, waulk; which happens to be ftill retained in fault. This was probably the ancient found of the Saxons, fince it is almost uniformly preferved in the rustic pronunciation, and the Northern dialects, as maun for man, haund for hand.
A open, not unlike the a of the Italians, is found in father, rather, and more obscurely in fancy, fast, &c.

The first letter of the European alphabets, has, in the English language, three different founds, which may be termed the broad, open, and slen-

The broad found refembling that of the German a is found, in many of our monofyllables, as all, wall, malt, falt; in which a is pronounced as au in caufe, or aw in law. Many of these words were anciently written with au, as fault, waulk; which happens to be still retained in fault. This was probably the ancient found of the Saxons, fince it is almost uniformly preferved in the rustic pronunciation, and the Northern dialects, as maun for man, haund for hand.
A open, not unlike the a of the Italians, is found in father, rather, and more obscurely in fancy, fast, &c.