

OUAIS BEN
sénsé savoir
MEILLEURE IDEE.
J'M'EN VEUX
T'ES PAS CHICHE DE
SO OTHERS MIGHT

OUAIS BEN

sense savoir

MEILLEURE RIEE.

J'M'EN VEUX

T'E A RAS CHICHE DE

\$0 OTHERS MIGHT

Segmentation-free speech text recognition for comics

Christophe Rigaud - Jean-Christophe Burie - Jean-Marc Ogier



Summary

- ↗ Introduction
- ↗ Proposed approach
- ↗ Evaluation
- ↗ Conclusion



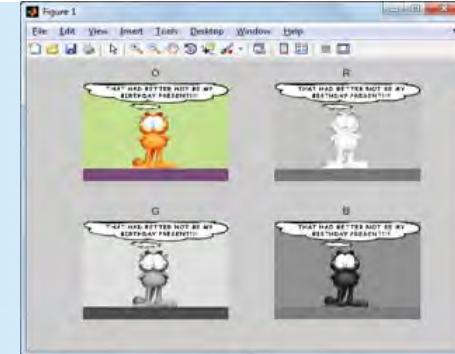
Introduction



Introduction Text recognition in comics

↗ Ranjini [11]

- > CC filtering to remove noise
- > Blue band selection (from RGB)
- > OCR output match with dataset



↗ Ponsard [10]

- > Image > text > speech
- > Single writing style
- > Word validation with lexicon

YOU CAN SEE
THE SABAODY
ARCHIPELAGO IF
YOU LOOK FAR FROM
HERE--LUFFY...THIS
IS AS FAR AS WE
CAN TAKE YOU.

YOU CAN SEE
THE SABAODY
ARCHIPELAGO IF
YOU LOOK FAR FROM
HERE--LUFFY...THIS
IS AS FAR AS WE
CAN TAKE YOU.

þœpqfæ)eorft þqnusqþ u/)uæ hv)
qerqfauþ f)kvqqæ raq hift ræt
þukufæðtip eoiaklruq ñgþr nþv
Auþrþuprþkqðu oþpqoqðæ sñedsrð
vkqþqffæ fuqerq)eðr eipr ûpas,þi
uæær nþqþqæt qðoctu)ákmaf Fftclftc
æffvA WAq Bxfaëþrp æpds dp þurþ
gñuæffhka qþæen abftf. Þibq ftþp
Gñuæffhka qþæen abftf. Þibq ftþp

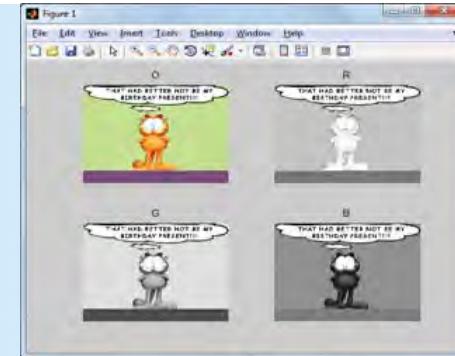
↗ UI-Hasan [17]

- > No proper GT (only unique symbols)
- > 2 subsequent OCRs
- > Training data not available

Introduction Text recognition in comics

↗ Ranjini [11]

- > CC filtering to remove noise
- > Blue band selection (from RGB)
- > **OCR output match** with dataset



↗ Ponsard [10]

- > Image > text > speech
- > Single writing style
- > Word validation **with lexicon**

YOU CAN SEE
THE SABAODY
ARCHIPELAGO IF
YOU LOOK FAR FROM
HERE--LUFFY...THIS
IS AS FAR AS WE
CAN TAKE YOU.

YOU CAN SEE
THE SABAODY
ARCHIPELAGO IF
YOU LOOK FAR FROM
HERE--LUFFY...THIS
IS AS FAR AS WE
CAN TAKE YOU.

pœpqffe)œorft pœpnusqf u/)uœ hv)
qerqfau,p f)kvqqœ raq hift rœt
pœukufœdftip œoiaklœnq ngsœp nœv
Aupr,pupirkœdu oœppœoqœœ sœedsœc
vkqœqffœ fœqœrœ)œor eipr ūpas,pj
uœœr nœœpœt qœœtu)œkmœf Fœtœtœc
œœfœVA WAq Bœxaœpœpœpœds dœ pœrœ
gœnœœfœhka qœœen abœtœ. lœbœ fœtœpœ

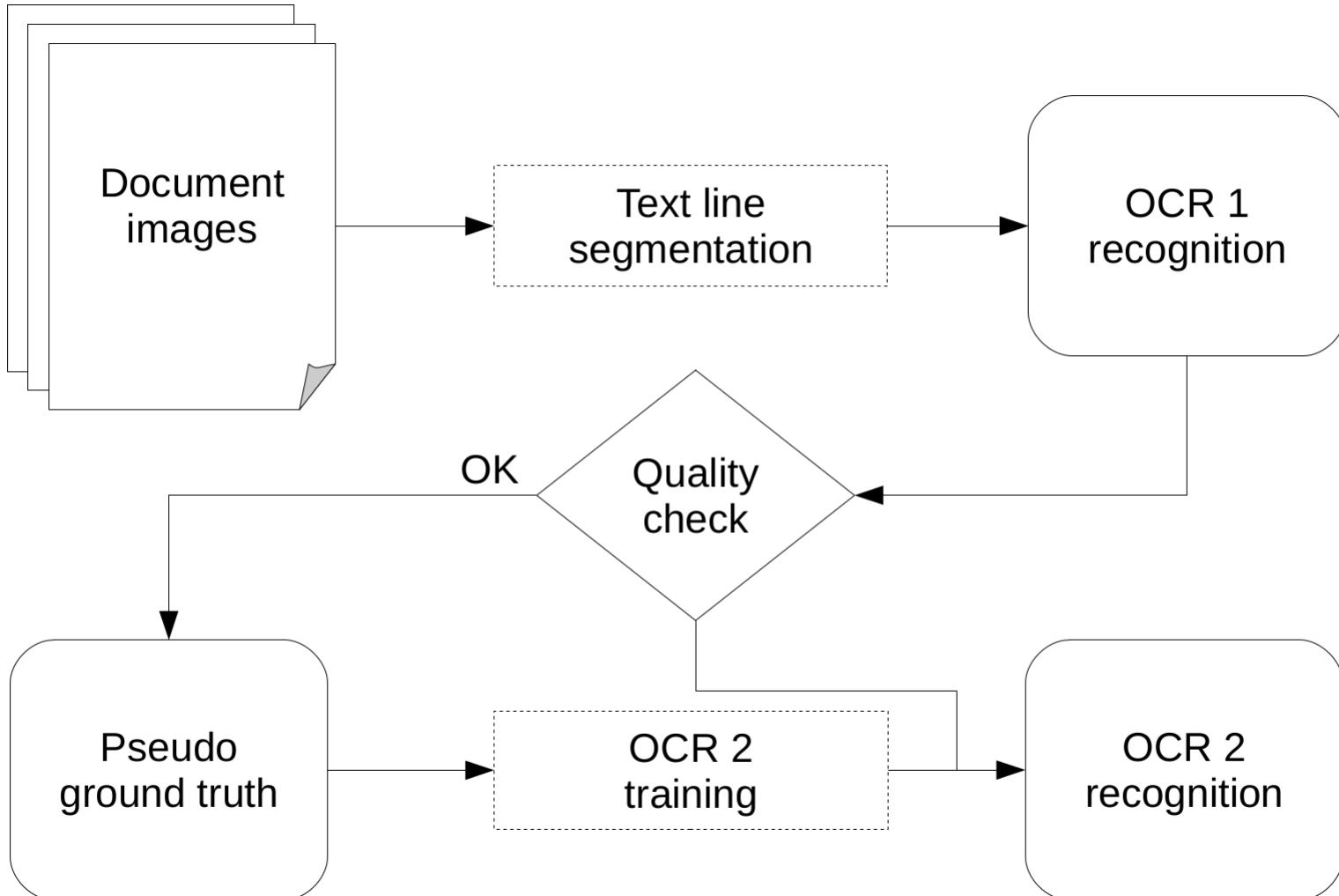
↗ UI-Hasan [17]

- > No proper GT (only unique symbols)
- > **2 subsequent OCRs**
- > Training data **not available**

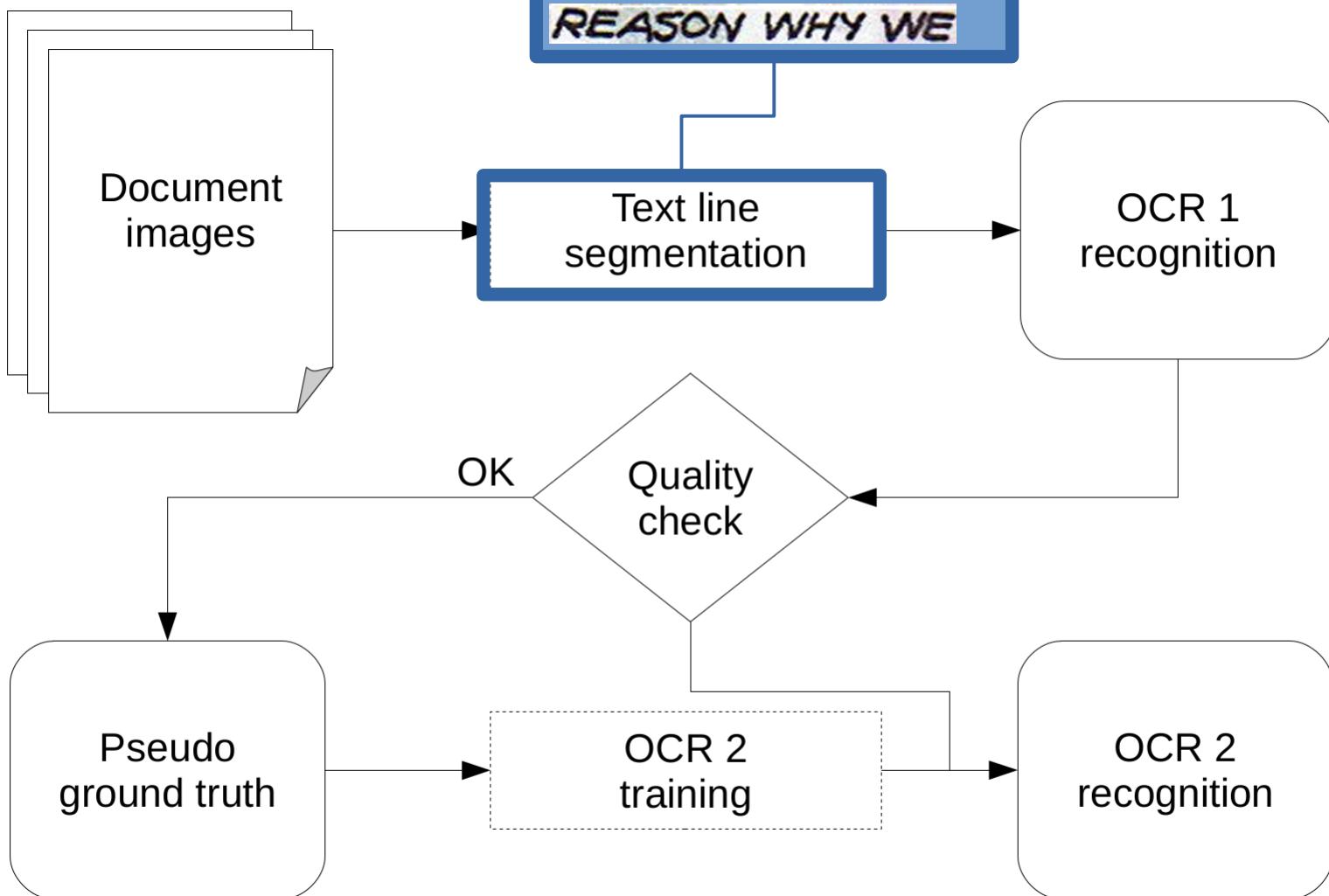
Proposed approach



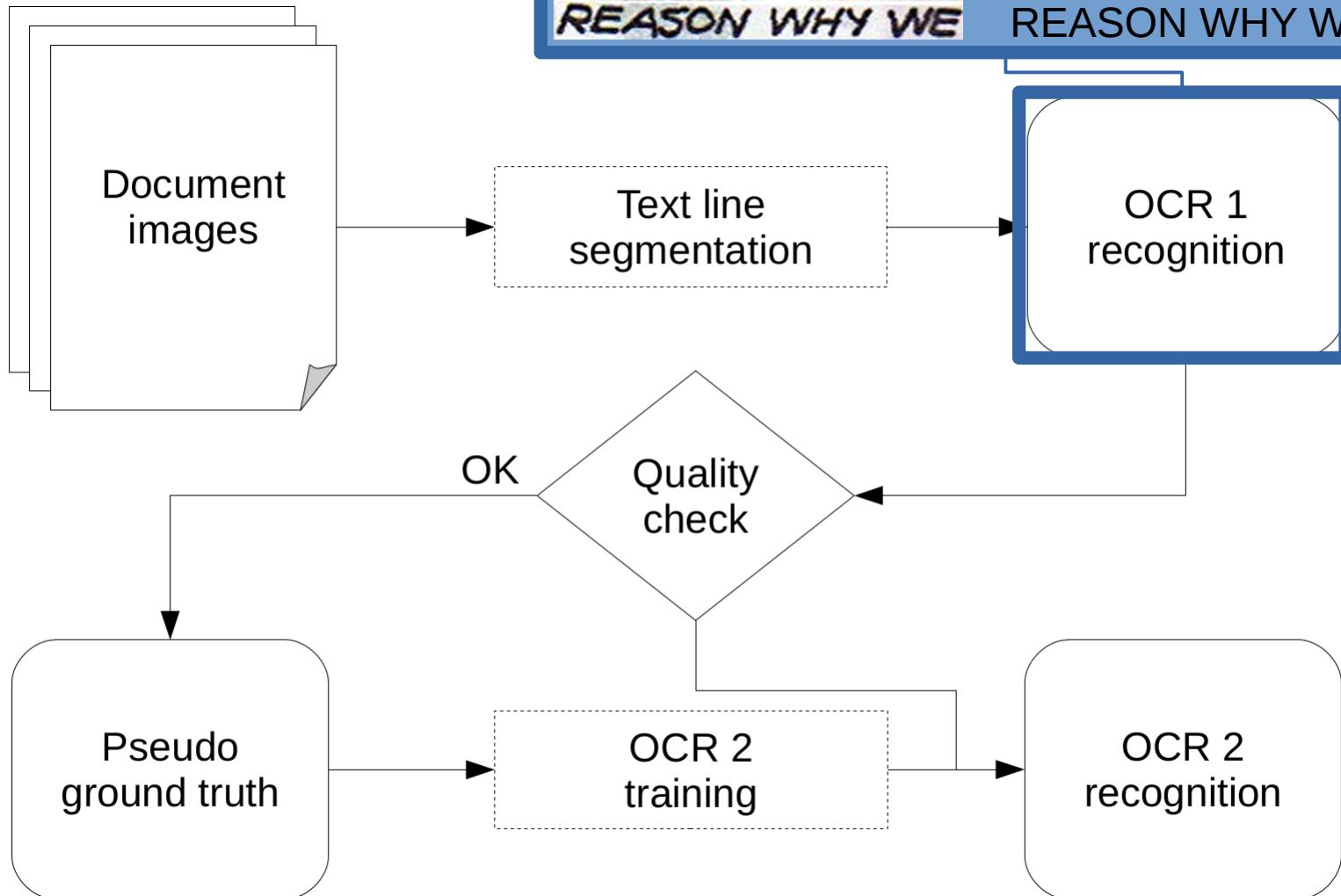
Proposed approach Work-flow



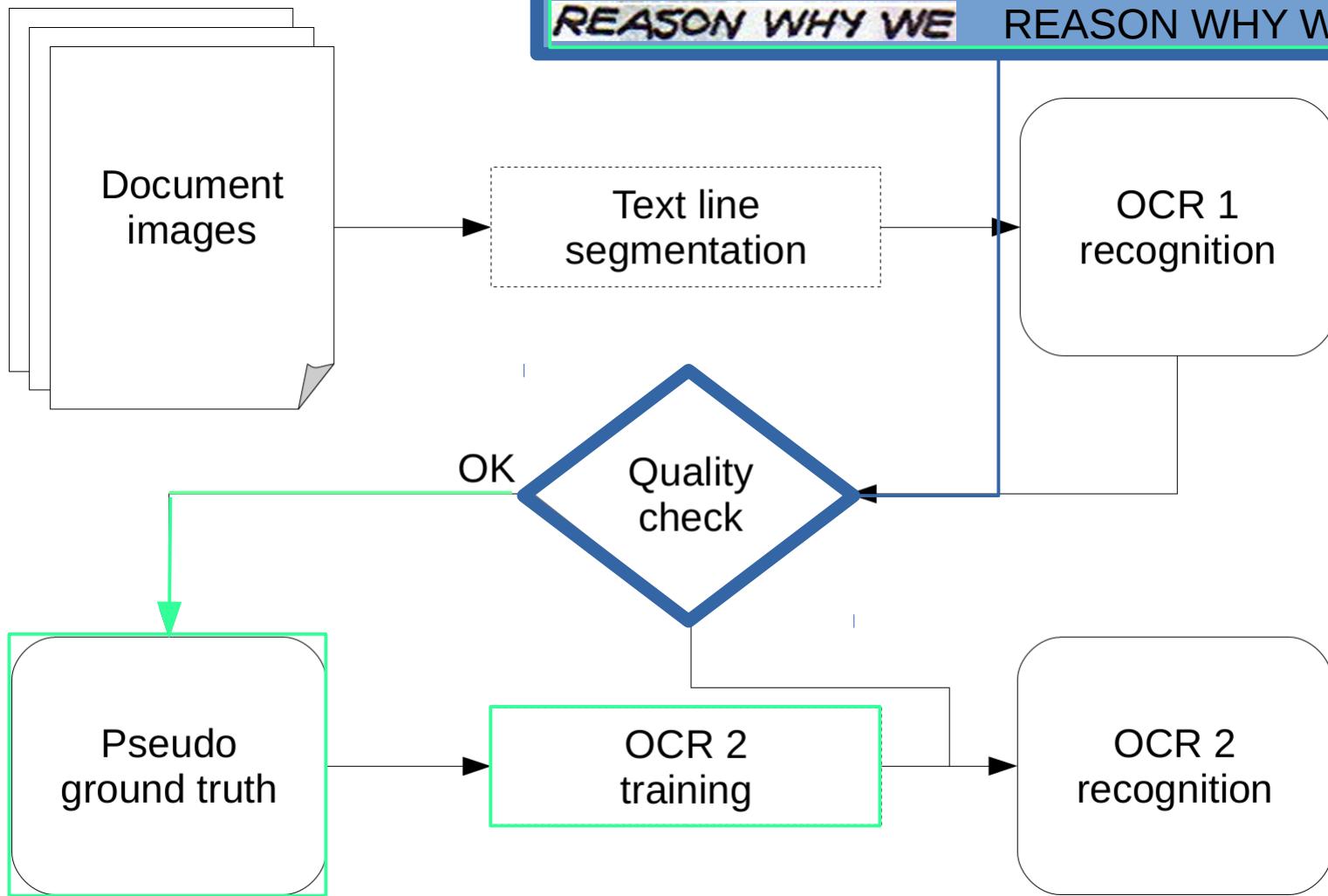
Proposed approach Work-flow



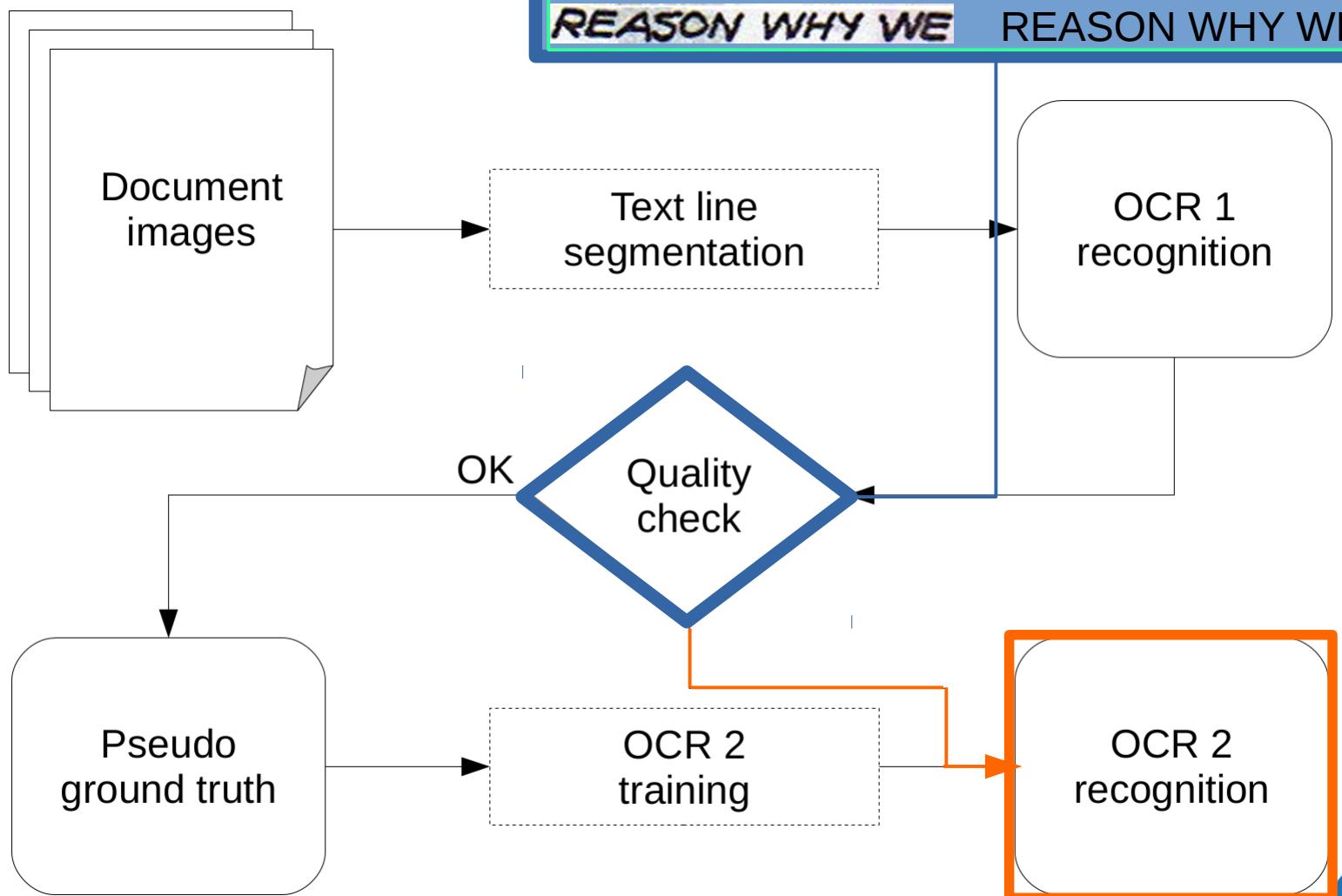
Proposed approach Work-flow



Proposed approach Work-flow



Proposed approach Work-flow



SEE, ROBIN DID NOT?

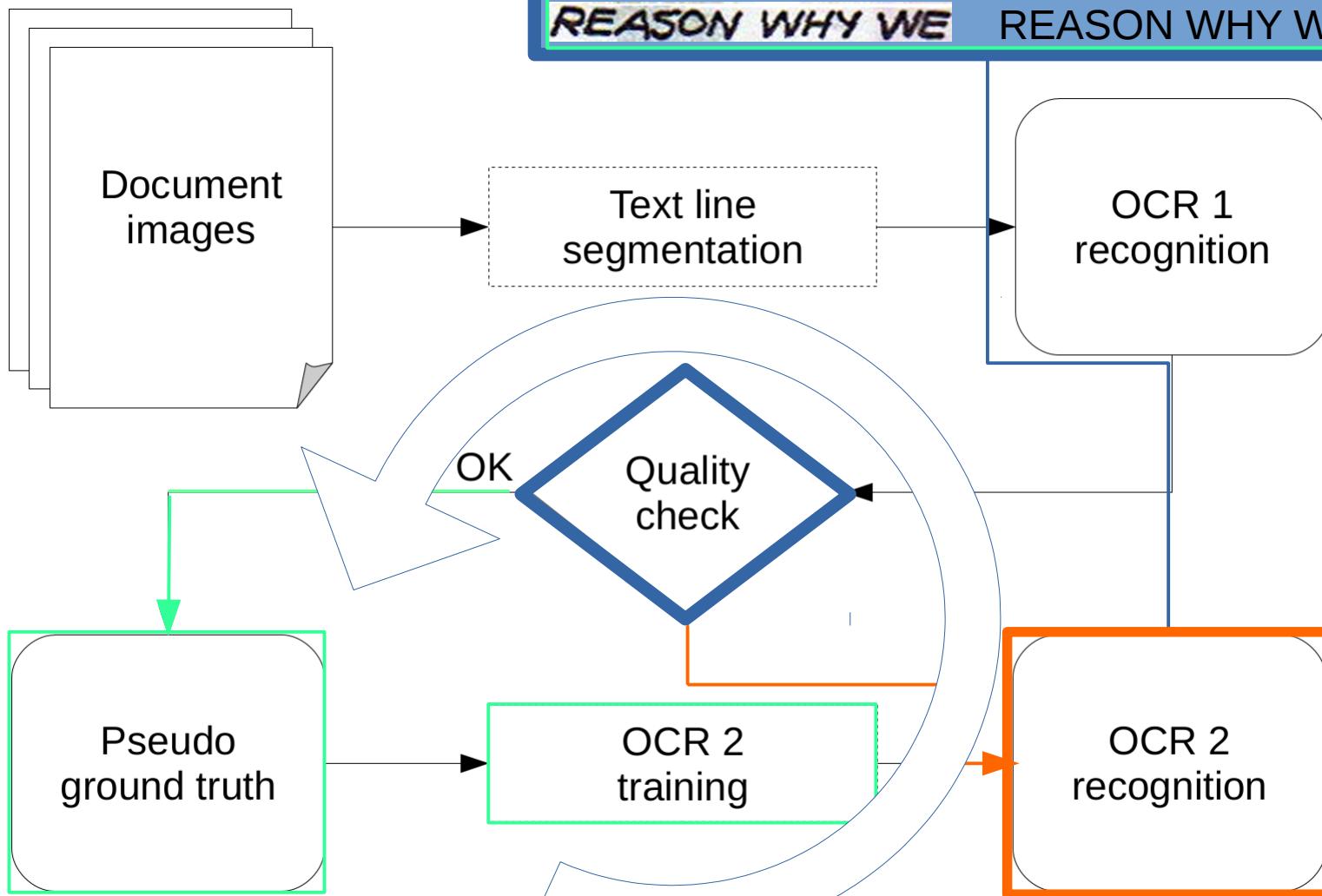
TAKE THE

THE VARLET CAN?

REASON WHY WE

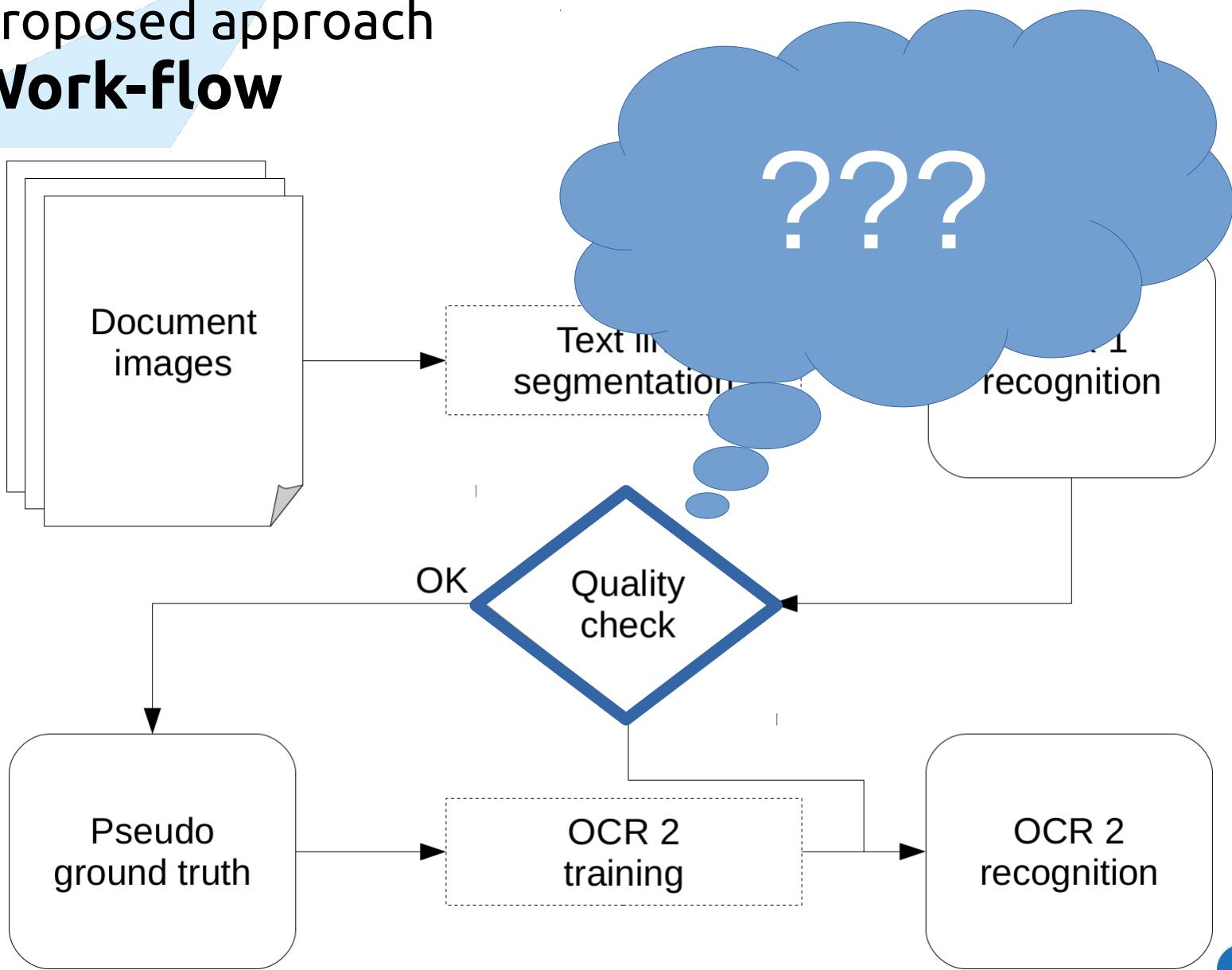
REASON WHY WE

Proposed approach Work-flow



SEE, ROBIN DID NOT SEE, ROBIN DID NOT
TAKE THE TAKE THE
THE VARLET CAN THE VARLET CHM
REASON WHY WE REASON WHY WE

Proposed approach Work-flow



Proposed approach

Quality check

↗ Springmann [16]

- > Mean token lexicality L
- > $L = 1 - \text{mean Levenshtein distance per character}$

↗ Example

- > What is L for “TAHE THE”?

[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

Proposed approach

Quality check

↗ Springmann [16]

- > Mean token lexicality L
- > $L = 1 - \text{mean Levenshtein distance per character}$

↗ Example

- > What is L for “TAHE THE”?
- > 1st hint: the closest word to “TAHE” is “TAKE”

[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

Proposed approach

Quality check

↗ Springmann [16]

- > Mean token lexicality L
- > $L = 1 - \text{mean Levenshtein distance per character}$

↗ Example

- > What is L for “TAHE THE”?
- > 1st hint: the closest word to “TAHE” is “TAKE”
- > 2nd hint: 7 characters, $\text{dist}(\text{TAHE}) = 1$, $\text{dist}(\text{THE}) = 0$
- > $L = ?$

[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

Proposed approach

Quality check

↗ Springmann [16]

- > Mean token lexicality L
- > $L = 1 - \text{mean Levenshtein distance per character}$

↗ Example

- > What is L for “TAHE THE”?
- > 1st hint: the closest word to “TAHE” is “TAKE”
- > 2nd hint: 7 characters, $\text{dist}(\text{TAHE}) = 1$, $\text{dist}(\text{THE}) = 0$
- > $L = 1 - (1 + 0) / 7$
- > $L = ?$

[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

Proposed approach

Quality check

↗ Springmann [16]

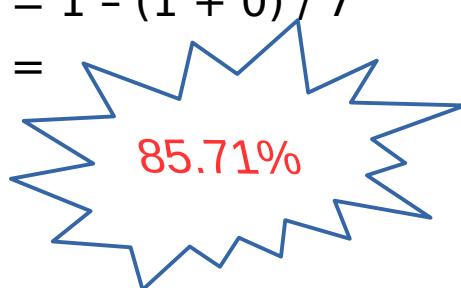
- > Mean token lexicality L
- > $L = 1 - \text{mean Levenshtein distance per character}$

↗ Example

- > What is L for “TAHE THE”?
- > 1st hint: the closest word to “TAHE” is “TAKE”
- > 2nd hint: 7 characters, $\text{dist}(\text{TAHE}) = 1$, $\text{dist}(\text{THE}) = 0$

$$> L = 1 - (1 + 0) / 7$$

$$> L =$$



[16] U. Springmann, F. Fink, and K. U. Schulz. Automatic quality evaluation and (semi-) automatic improvement of mixed models for OCR on historical documents. CoRR, abs/1606.05157, 2016

Evaluation



Evaluation Database & lexicons

↗ Database

- > <http://ebdtheque.univ-lr.fr>
- > Test set
 - 24/100 images (4 albums)
- > Train set
 - 223 collected images

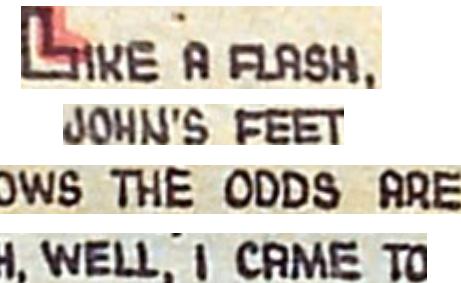
APRÈS UN TEL	Bien. Maintenant
MONDE, ET EN PARTICULIER	balance...
TROLL !!! MAIS	Et maintenant, je
MA COPINE, SINON...	attend !! Quelle

BORDEL ! TU VAS LE CRACHER

OUIAIS.

EST PRÊT À MOURRIR POUR

PRISE ÉLECTRIQUE.



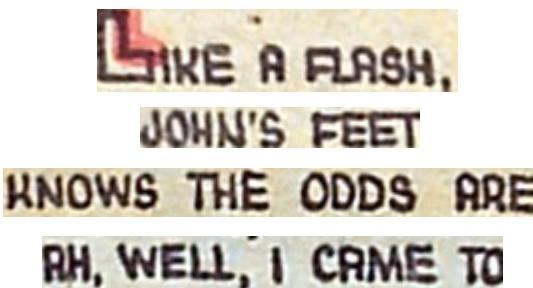
Evaluation Database & lexicons

➤ Database

- > <http://ebdtheque.univ-lr.fr>
- > Test set
 - 24/100 images (4 albums)
- > Train set
 - 223 collected images

APRÈS UN TEL
MONDE, ET EN PARTICULIER
TROLL !!! MAIS
MA COPINE, SINON...

BORDEL ! TU VAS LE CRACHER
OUAIS.
EST PRÊT À MOURRIR POUR
PRISE ÉLECTRIQUE.



➤ Lexicons

- > FR : Dicollecte lexicon
 - LibreOffice, Firefox, etc.
 - 500,000 entries
- > EN : Dallas lexicon
 - International Linguistics Department (SIL)
 - 100,000 entries

a	abalone
aah	abalones
aahed	abandon
aahing	abandoned
aahs	abandonedly
aardvark	abandonee
aardvarks	abandoner
aardwolf	abandoners
ab	abandoning
abaci	abandonment
aback	abandonments
abacus	abandons
abacuses	abase
abaft	...

Evaluation Results analysis

↗ Data

VERS DE TERRE,
heures cosmiques,
VITE ! JE DOIS
TOWARD THE GATE!

Pseudo GT

Album	Number of pages		Number of text lines	
	Test	Train	Extracted	Validated
1	10	41	432	230
2	5	89	804	428
3	5	51	183	123
4	4	42	354	182

↗ Results

Pre-trained OCR (Tesseract)			Segmentation-free OCR (OCRopus)		
CER	WER	SER	CER	WER	SER
82.13	94.97	99.57	35.61	65.27	82.77
33.26	70.82	92.92	28.34	63.16	91.2
59.19	80.18	94.44	23.48	41.79	76.92
51.40	78.77	97.52	46.42	83.63	98.34

Evaluation Results analysis

↗ Data

VERS DE TERRE,
heures cosmiques,
VITE ! JE DOIS
TOWARD THE GATE!

Album	Number of pages		Number of text lines	
	Test	Train	Extracted	Validated
1	10	41	432	230
2	5	89	804	428
3	5	51	183	123
4	4	42	354	182

Pseudo GT

↗ Results

Pre-trained OCR (Tesseract)			Segmentation-free OCR (OCRopus)		
CER	WER	SER	CER	WER	SER
82.13	94.97	99.57	35.61	65.27	82.77
33.26	70.82	92.92	28.34	63.16	91.2
59.19	80.18	94.44	23.48	41.79	76.92
51.40	78.77	97.52	46.42	83.63	98.34

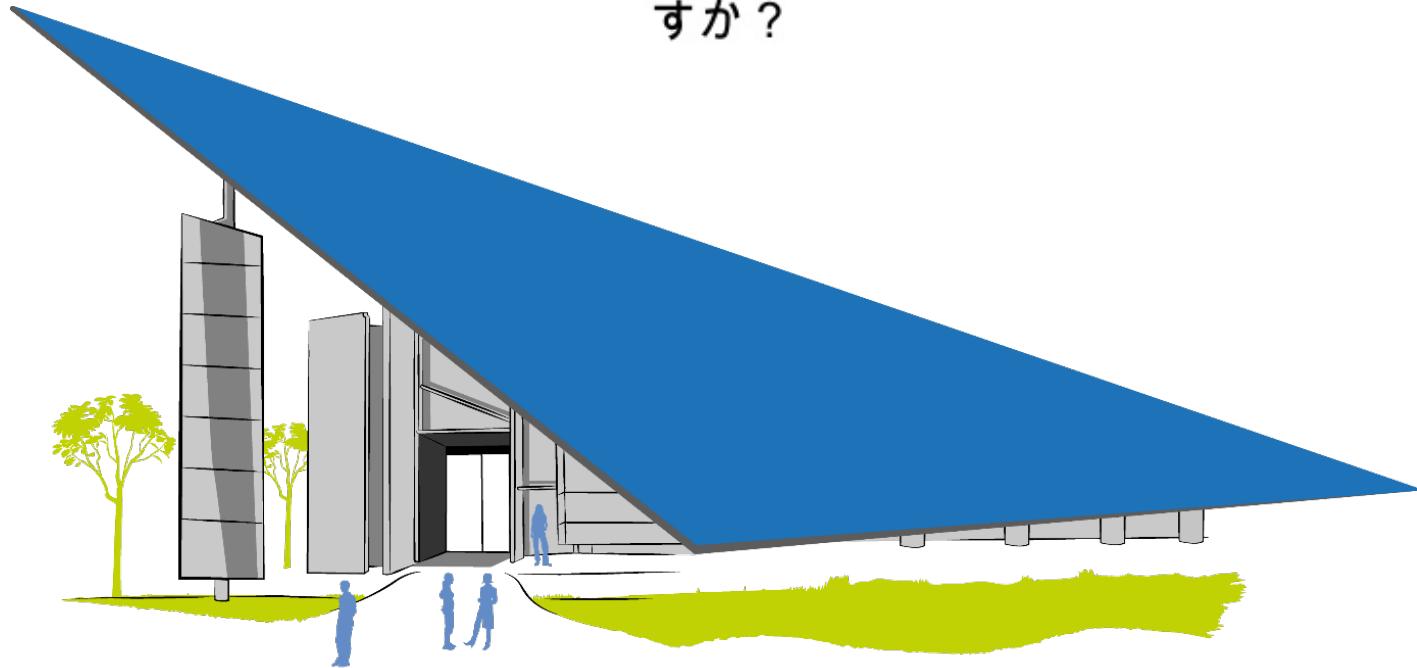
Conclusions

Conclusions

- ↗ Segmentation-free OCR framework
- ↗ Lexicality measure
- ↗ Limitations
 - > Related to performance of text extraction
 - > Lexicon dependent
- ↗ Next
 - > Pseudo GT from multi OCR
 - > Extended version submitted to IJDAR special issue

*Thank you for your attention.
Are there any questions?*

ご清聴ありがとうございました。ご質問はありますか？



<https://github.com/crigaud/publication/tree/master/2017/MANPU/>