

# Workshop on Document Intelligence Understanding

<https://doc-iiu.github.io/>

## ABSTRACT

Document understanding and information extraction include different tasks to understand a document and extract valuable information automatically. Recently, there has been a rising demand for developing document understanding among different domains, including business, law, and medicine, to boost the efficiency of work that is associated with a large number of documents.

This workshop aims to bring together researchers and industry developers in the field of document intelligence and understanding diverse document types to boost automatic document processing and understanding techniques. We also release a data challenge on the recently introduced document-level VQA dataset, PDFVQA<sup>1</sup>. The PDFVQA challenge examines the model's structural and contextual understandings on the natural full document level of multiple consecutive document pages by including questions with a sequence of answers extracted from multi-pages of the full document. This task helps to boost the document understanding step from the single-page level to the full document level understanding.

## KEYWORDS

Document Understanding, Information Extraction, Layout Analyzing, Visual Question Answering

### ACM Reference Format:

Workshop on Document Intelligence Understanding <https://doc-iiu.github.io/>. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 4 pages. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 WORKSHOP ORGANIZERS

### Dr Caren Han

The University of Sydney and The University of Western Australia, Australia, [caren.han@sydney.edu.au](mailto:caren.han@sydney.edu.au) and [caren.han@uwa.edu.au](mailto:caren.han@uwa.edu.au)

### Mr. Yihao Ding

The University of Sydney, Australia, [yihao.ding@sydney.edu.au](mailto:yihao.ding@sydney.edu.au)

### Ms. Siwen Luo

The University of Sydney and The University of Western Australia, Australia, [siwen.luo@sydney.edu.au](mailto:siwen.luo@sydney.edu.au) and [siwen.luo@uwa.edu.au](mailto:siwen.luo@uwa.edu.au)

### Dr. Josiah Poon

The University of Sydney, Australia, [josiah.poon@sydney.edu.au](mailto:josiah.poon@sydney.edu.au)

<sup>1</sup><https://www.kaggle.com/t/ce5b8d5610c24d719d9a76020700f8bf>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00  
<https://doi.org/XXXXXXX.XXXXXXX>

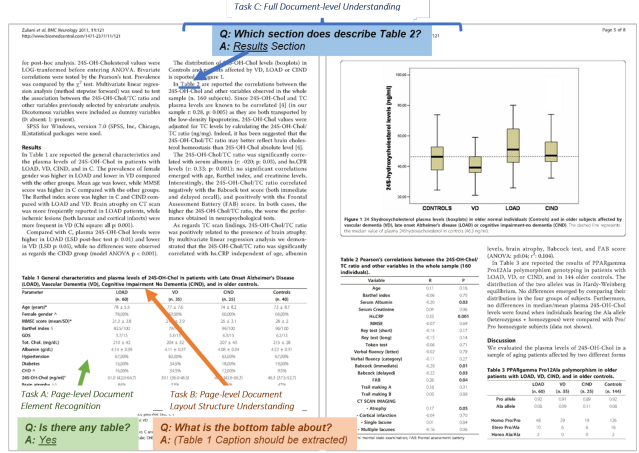


Figure 1: PDF-VQA Data Challenge Sample Questions and Document Pages for Task A, B, and C. [2]

## 2 WORKSHOP CONTACT PERSON

### Dr. Caren Han

Address: The University of Western Australia (M002), 35 Stirling Highway, 6009 Perth, Australia (+61 423915170)

Email: [caren.han@uwa.edu.au](mailto:caren.han@uwa.edu.au)

Website: <https://drcarenhan.github.io/>

## 3 WORKSHOP THEME AND TOPICS

This workshop aims to explore and advance the current state of research and technical articles, including but not limited to the topics. Note that the models for the data challenges should apply the following technology:

**Document Processing and Structure Understanding:** Benchmark datasets, models, and off-the-shelf applications for processing unstructured or semi-structured document images into machine-readable formats, including but not limited to:

- Document Layout Analysis
- Document Parsing
- Document Image Processing
- Document Table Detection
- Table Structure Recognition
- Reading Order Prediction

**Document Content Understanding:** Benchmark datasets, models and off-the-shelf applications focusing on understanding the content of target documents to extract critical information and conducting downstream analysis:

- Document Visual Question Answering
- Document Information Retrieval
- Document Key Information Extraction
- Document Classification and Categorization

Workshop Name	Conference	Year	Website
<b>Document Intelligence or Processing Workshops</b>			
DocVQA	ICDAR	2021	<a href="https://www.docvqa.org/workshops">https://www.docvqa.org/workshops</a>
Scholarly Document Processing	EMNLP, NAACL, COLING	2020, 2021, 2022	<a href="https://sdproc.org/2022/index.html">https://sdproc.org/2022/index.html</a>
Document Intelligence	KDD	2022	<a href="https://document-intelligence.github.io/DI-2022/">https://document-intelligence.github.io/DI-2022/</a>
<b>Information Retrieval Workshops (including document understanding aspect)</b>			
Proactive And Agent-Supported Information Retrieval	CIKM	2022	<a href="https://pasircikm2022.github.io/PASIRCIKM/">https://pasircikm2022.github.io/PASIRCIKM/</a>
Federated Learning for Information Retrieval	SIGIR	2023	<a href="https://sites.google.com/view/flirt-sigir23">https://sites.google.com/view/flirt-sigir23</a>

Table 1: Related Workshop List in Previous Conferences.

- Historical Document Content Understanding
- Table-based Question Answering

**Advanced Deep Learning Approaches for Doc-IU:** Recently proposed deep learning techniques boosting the development of document intelligence including but not limited to:

- Document Feature Representation
- Multi-modal Fusion and Adaptive Learning
- Pre-training Mechanisms for Document Understanding
- Few-shot Learning and Zero-shot Learning

#### 4 WORKSHOP OBJECTIVES, GOALS & EXPECTED OUTCOME

This workshop aims to invite and share recent research works and technical reports of various document understanding tasks, including document layout analysis[1, 4, 5, 9], document visual question answering[2, 6, 8], document key-value extraction[3, 7]. It provides a discussion panel for researchers in this field to share and discuss the research work trend and techniques for better document understanding.

Moreover, this workshop proposes a document VQA data challenge with the PDFVQA dataset (Challenge website: <https://www.kaggle.com/competitions/pdfvqa>). This challenge aims to develop models to answer the questions with the given document images. Unlike the other document VQA datasets that focus on the contextual understanding of texts, PDFVQA questions target the structural relationship understandings among the document layout components. Moreover, instead of processing on independent document images, PDFVQA datasets contain the whole document of multiple pages, and there could be multiple answers that span over multiple pages for one question. PDFVQA motivates the document understanding and processing not limited to the single document page but expands to the full document level to understand the logic and connections of document contents and structures over consecutive pages. We expect through the proposed challenge, researchers in document understanding would pay more attention to and develop new techniques for the document understanding task on the full document level.

As shown in Figure 1, answering a question requires reviewing the full document contents and identifying the contents hierarchically related to the queried item in the question. For example, the question “Which section does describe Table 2?” in Figure 1 requires the identification of all the sections of the full document that have

described the queried table. The answers to such questions are the texts of the corresponding section titles extracted as the high-level summarization of the identified sections. Identifying the items at the higher-level hierarchy of the queried item is defined as the parent relation understanding the question in PDF-VQA. Oppositely, Task C also contains the questions of identifying the items at the lower-level hierarchy of the queried item, and such questions are defined as the child relation understanding. For example, a question, “What does the ‘Methods’ section about?” requires extracting all the subsection titles as the answer.

#### 5 WORKSHOP LENGTH

The workshop will be hosted in half-day (4 hours<sup>2</sup>), containing three main sessions. Each session is scheduled to be roughly 1 hour (60 mins), with two 15-min breaks in between. The PDF dataset challenge session would include the Winners’ presentation and the Award Ceremony. The Award Ceremony will be sponsored by Google Research and FortifyEdge. The detailed length and schedule are in Table 2.

#### 6 TARGET AUDIENCE

The workshop would be beneficial to researchers, industrial developers, and practitioners who are interested in broad information retrieval, knowledge management, natural language processing, machine learning, and data mining, specifically document understanding, document intelligence, and document content understanding. While the audience with a good background in the above areas would benefit most from this workshop, we believe that the presented research keynote speech, and model architectures and technical details to address the dataset challenge would give the general audience and newcomers a complete picture of the current work and inspire them to learn more in this field. The workshop is designed as self-contained, so no specific background knowledge is assumed of the audience. However, it would be advantageous for the audience to know about basic multi-modal deep learning technologies and new applications of large pre-trained models on documents (a new type of image source). We will provide the audience with the reading list and dataset samples on our workshop website. Note that the workshop website is provided in <https://doc-iiu.github.io/>.

<sup>2</sup>This can be updated based on the decision made by the CIKM workshop chairs

Length	Invite Speaker	Topic
10 mins	Dr. Caren Han	Workshop Opening Speech
<b>Document Understanding Datasets Session</b>		
60 mins		Document Understanding Benchmark Dataset papers published in CIKM, SIGIR, KDD, ICDM, ACL, EMNLP and NAACL conferences 2021-2023
15 mins		<b>break</b>
<b>Document Understanding Model Session</b>		
60 mins		Document Understanding Model papers published in top-tier CIKM, SIGIR, KDD, ICDM, ACL, EMNLP and NAACL conferences 2021-2023
15 mins		<b>break</b>
<b>PDFVQA Challenge Session</b>		
10 mins	Mr. Yihao Ding	Leaderboard Opening
20 mins	Ms. Siwen Luo	PDF-VQA: A New Dataset for Real-World VQA on PDF Documents
30 mins	Winner	Challenge Winner Presentation on Methodology and Results
10 mins	Dr. Josiah Poon	Awards Ceremony
10 mins	Dr. Caren Han	Conclusion Remarks

Table 2: Workshop Program Schedule

## 7 WORKSHOP RELEVANCE

The workshop is highly relevant to the CIKM community on research on information and knowledge management. It also directly aligned with the aim of the CIKM workshops, bridging the academic-commercial gap in the database, information retrieval, machine learning, and knowledge management communities. As informed in the CIKM workshop proposal website, the proposed workshop, DocIU, would enable participants to propose novel research and present practical applications on document intelligence and understanding by addressing the research problem of the dataset challenge, PDF-VQA. It would require many aspects of the data lifecycle, including acquisition, pre-processing, modelling, integration/aggregation, and analysis. The workshop finally related to the aim of the CIKM workshop, providing interdisciplinary workshops bridging across different communities. The aim of the proposed workshop and workshop organizers have rich interdisciplinary aspects in major Natural Language Processing, Information Retrieval, Computer Vision, and Knowledge Management.

## 8 RELATED WORKSHOPS

The workshop is considered a cutting-edge workshop that covers the recent trends in emerging areas of information retrieval, knowledge management, and natural language processing. As shown in Table 1, previous workshops in top-tier Natural Language Processing conferences, including EMNLP, NAACL, COLING, and KDD, covered similar but not the same topics in document intelligence and processing. The DocVQA workshop was only limited to the Document Visual Question Answering, while the scholarly Document Processing workshop has a broader document processing topic range, not focusing on the multimodal document image and text processing as our proposed workshop. Their topics focus more

on text processing, including generation and summarization. Document Intelligence workshop focuses more on document understanding tasks. But their topics are limited to business document understanding while our workshop encourages broader exploration of diverse document types.

The CIKM or SIGIR workshops have covered information retrieval or federated learning using document types but have not touched on the exact document understanding or intelligence in the past three years. The workshop has not been released elsewhere.

## 9 WORKSHOP PROGRAM FORMAT

Workshop chairs Dr. Caren Han and Ms. Siwen Luo and Leaderboard Chair Mr. Yihao Ding will organize the workshop in person. The workshop contains three main sessions. The first session includes the works to introduce different document understanding tasks and the benchmark dataset works. The second session includes the works of key document understanding models. The last session will be the PDFVQA challenge session. The challenge winner will present the methodology and results, followed by a QA session. The detailed workshop program is listed in Table 2.

## 10 WORKSHOP SCHEDULE & IMPORTANT DATES

- Leaderboard Challenge Due: 15 September 2023
- Workshop Paper Submission Due: 15 September 2023
- Announcement of Winner: 30 September 2023
- Paper Acceptance notification: 30 September 2023
- Workshop Date: 22 October 2023

## 11 PROGRAM COMMITTEE

- Workshop Chair: Dr. Caren Han, The University of Sydney & The University of Western Australia, Australia

- Workshop Chair: Mr. Yihao Ding, The University of Sydney, Australia
- Workshop Chair: Ms. Siwen Luo, The University of Sydney & The University of Western Australia, Australia
- Workshop Chair: Dr. Josiah Poon, The University of Western Australia, Australia
- Advisory Committee: Mr. Zhe Huang, ANT Group & Alibaba Group, China
- Advisory Committee: Dr. HeeGuen Yoon, National Information Society Agency, Korea
- Advisory Committee: Dr. Paul Duuring, Department of Mines and Petroleum, Australia
- Advisory Committee: Prof. Eun-Jung Holden, UWA Institute of Data, Australia

## 12 PARTICIPATION & SELECTION PROCESS

Participants are selected with their works in the document understanding domain published at top-tier Information Retrieval and Natural Language Processing conferences, including CIKM, SIGIR, KDD, ICDM, ACL, EMNLP and NAACL.

For the PDFVQA challenge, the winner will be chosen via the top leaderboard score. In addition, challenge participants must submit all technical reports, including the code and a 2-page report illustrating the methodology and testing results. The winner selection criteria also include the validity and reproductivity of the submitted code and the quality of the report.<sup>3</sup>

## 13 ORGANIZERS' BACKGROUND

The workshop organizers' are all experts in Natural Language Processing and Information Retrieval, and the detailed background information can be found as follows:

**Mr. Yihao Ding (In person)** is a PhD candidate at the School of Computer Science, University of Sydney, and visiting scholar at the School of Computer Science, University of Western Australia. He is a Research Assistant at the School of Earth Science, University of Western Australia. He received his Bachelor's Degree and Master's Degree in 2015 and 2018 in Geospatial Engineering and his second Master's Degree in Information Technology in 2020. His research interests include deep learning-based document analysis, information retrieval, and graph neural networks. He has published several top-tier conferences and journal papers in CVPR, SIGIR, COLING, ECML-PKDD, and Frontiers.

**Ms. Siwen Luo (In person)** is a final year PhD student at the School of Computer Science, The University of Sydney, and a visiting scholar at the School of Physics, Mathematics and Computers, University of Western Australia. Her research focuses on the cross-area of computer vision and natural language processing, aiming for the exploration and development of interpretable models for multimodalities. Her research spans a range of multimodal tasks, including Visual Question Answering, Text to Image generation, and Document Layout Analysis. She has published several papers at top-tier NLP conferences and received the Best Paper Award from ICONIP 2020 and the best area paper from COLING 2020.

**Dr. Soyeon Caren Han (In person)** is a co-leader of AD-NLP (Australia Deep Learning NLP Group) and a Senior lecturer (Associate Professor in U.S. System) at the University of Western Australia and an honorary senior lecturer (honorary Associate Professor in U.S. System) at the University of Sydney and the University of Edinburgh. After her Ph.D.(in 2017), she worked for six years at the University of Sydney. Her research interests include Natural Language Processing with Deep Learning. She is broadly interested in several research topics, including visual-linguistic multimodal learning, abusive language detection, document analysis, and recommender system. More information can be found at <https://drcarenhan.github.io/>.

**Dr. Josiah Poon** is a co-leader of AD-NLP (Australia Deep Learning NLP Group) and a Senior Lecturer at the School of Computer Science, University of Sydney. He's been using traditional machine learning techniques paying particular attention to learning from imbalanced datasets, short string text classification, and data complexity analysis. He has coordinated a multidisciplinary team consisting of computer scientists, pharmacists, western medicine & traditional Chinese medicine researchers and practitioners since 2007. He co-leads a joint big-data laboratory for integrative medicine (Acclaim) established between the University of Sydney and the Chinese University of Hong Kong to study medical/health problems using computational tools.

## 13.1 Advisory Committees

We also invite four advisory committees from diverse domain backgrounds and different countries:

**Dr. HeeGuen Yoon**, National Information Society Agency, **Korea**  
**Mr. Zhe Huang**, ANT Group, Alibaba Group, **China**  
**Dr. Paul Duuring**, Department of Mines and Petroleum, **Australia**  
**Prof. Eun-Jung Holden**, UWA Institute of Data, **Australia**

## REFERENCES

- [1] Yihao Ding, Zhe Huang, Runlin Wang, YanHang Zhang, Xianru Chen, Yuzhong Ma, Hyunsuk Chung, and Soyeon Caren Han. 2022. V-Doc: Visual questions answers with Documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 21492–21498.
- [2] Yihao Ding, Siwen Luo, Hyunsuk Chung, and Soyeon Caren Han. 2023. PDFVQA: A New Dataset for Real-World VQA on PDF Documents. *arXiv preprint arXiv:2304.06447* (2023).
- [3] Zheng Huang, Kai Chen, Jianhua He, Xiang Bai, Dimosthenis Karatzas, Shijian Lu, and CV Jawahar. 2019. Icdar2019 competition on scanned receipt ocr and information extraction. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1516–1520.
- [4] Minghao Li, Yiheng Xu, Lei Cui, Shaohan Huang, Furu Wei, Zhoujun Li, and Ming Zhou. 2020. DocBank: A Benchmark Dataset for Document Layout Analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*. 949–960.
- [5] Siwen Luo, Yihao Ding, Siqu Long, Josiah Poon, and Soyeon Caren Han. 2022. Doc-GCN: Heterogeneous Graph Convolutional Networks for Document Layout Analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*. 2906–2916.
- [6] Minesh Mathew, Dimosthenis Karatzas, and CV Jawahar. 2021. Docvqa: A dataset for vqa on document images. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2200–2209.
- [7] Seunghyun Park, Seung Shin, Bado Lee, Junyeop Lee, Jaeheung Surh, Minjoon Seo, and Hwalsuk Lee. 2019. CORD: a consolidated receipt dataset for post-OCR parsing. In *Workshop on Document Intelligence at NeurIPS 2019*.
- [8] Ryota Tanaka, Kyosuke Nishida, and Sen Yoshida. 2021. Visualmrc: Machine reading comprehension on document images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 13878–13888.
- [9] Xu Zhong, Jianbin Tang, and Antonio Jimeno Yepes. 2019. Publaynet: largest dataset ever for document layout analysis. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 1015–1022.

<sup>3</sup>Due to the Award Prize transfer, note that attendees from Countries in Sanctions list will not be considered