Short Paper Booklet

15th IAPR International Workshop on Document Analysis Systems

22 — 25 May 2022

DAS 2022 - La Rochelle

La Rochelle Université

L3i Informatique Image Interaction La Rochelle Univ.

IAPR

FOREWORD

We are very happy to welcome you to DAS 2022, the 15th IAPR International Workshop on Document Analysis Systems, held in La Rochelle, France, for the first time. Organizing an international workshop of such significant size after the COVID pandemic, aiming to welcome most of the participants on-site, is a challenge we are very happy to have taken on. Defining best-practice in organizing large hybrid events remains an on-going effort for the scientific community and we hope to have ensured a pleasant experience both for on- and off-site participants.

At the time of writing, over 70% of the registrations are for on-site participation. We are looking forward to hosting our friends and colleagues of the DAS community 4 years after we could last meet face-to-face. We are especially pleased to provide this opportunity to young researchers, some of whom will attend their first ever in-person scientific event. We supported their participation with considerably reduced registration fees for students and a financial assistance program.

We hope you will enjoy our city of La Rochelle. Located on the Atlantic coast of France, La Rochelle has recently been ranked among the most livable cities in France, in particular for students. The city has a rich historical fabric, with its old harbor and towers as its most well-known landmarks.

We will treat you the best way possible with a welcome cocktail in a splendid 18th century cloister (*Cloître des Dames Blanches*), part of the city hall, and a gala dinner in the old harbor, preceded by a sea tour to the picturesque *Fort Boyard* (as seen on TV in 70 countries).

The workshop will be hosted on-campus by La Rochelle Université, using state-of-the-art broadcasting equipment. The campus and all its workshop venues are located within walking distance of the conference hotels, the historic center and the *Minimes* beach.

Finally, we want to thank the numerous and deeply committed volunteers of the local organization team. Without them, and the support of this year's workshop sponsors, this 15th IAPR International Workshop on Document Analysis Systems would not be possible. Last but not least, we want to thank you, the participants, our friends and colleagues, for giving us the pleasure of your attendance, whether online or offline.

Welcome to La Rochelle!

May 2022

Jean-Marc OGIER,
Jean-Christophe BURIE
Mickaël COUSTATY
Antoine DOUCET

PREFACE

Welcome to the 15th IAPR International Workshop on Document Analysis Systems (DAS 2022). DAS 2022 was held in La Rochelle, France, during May 22–25, 2022, and brought together many researchers from Europe and abroad.

With the new remote access facilities, the workshop was not confined to a specific location. In a sense, this was truly a worldwide edition of DAS, taking place around the world in a coordinated fashion, employing a schedule we designed to support participation across a wide range of time zones. Of course, this came with some challenges, but also with interesting opportunities that caused us to rethink the way of fostering social and scientific interaction in this new medium. It also allowed us to organize an environmentally friendly event, extend the reach of the workshop, and facilitate participation from literally anywhere in the world for those with an interest in our field and an Internet connection. We truly hope we managed to make the most out of a difficult situation.

DAS 2022 continued the long tradition of bringing together researchers, academics, and practitioners in the research field of document analysis systems. In doing so, we built upon the previous workshops held over the years in Kaiserslautern, Germany (1994); Malvern, PA, USA (1996); Nagano, Japan (1998); Rio de Janeiro, Brazil (2000); Princeton, NJ, USA (2002); Florence, Italy (2004); Nelson, New Zealand (2006); Nara, Japan (2008); Boston, MA, USA (2010); Gold Coast, Australia (2012); Tours, France (2014); Santorini, Greece (2016); Wien, Austria (2018); and Wuhan, China (2020).

As with previous editions, DAS 2022 was a rigorously peer-reviewed and 100% participation single-track workshop focusing on issues and approaches in document analysis and recognition. The workshop comprised presentations by invited speakers, oral and poster sessions, and a pre-workshop tutorial, as well as distinctive DAS discussion groups.

This year we received 94 submissions in total, 78 of which were in the regular paper track and 16 in the short paper track. All regular paper submissions underwent a rigorous single-blind review process where the vast majority of papers received three reviews. The reviewers were selected from the 80 members of the Program Committee, judging the originality of work, the relevance to document analysis systems, the quality of the research or analysis, and the overall presentation. Of the 78 regular submissions received, 52 were accepted for presentation at the workshop (67%). Of these, 31 papers were designated for oral presentation (40%) and 21 for poster presentation (27%). All short paper submissions were reviewed by all three program co-chairs. Of the 16 short papers received, all 16 were accepted for poster presentation at the workshop (100%). The accepted regular papers are published in this proceedings volume in the Springer Lecture Notes in Computer Science series. Short papers appear in PDF form on the DAS conference website.

The final program included six oral sessions, two poster sessions, and the discussion group sessions. There were also two awards announced at the conclusion of the workshop: the IAPR Best Student Paper Award and the IAPR Nakano Best Paper Award. We offer our deepest thanks to all who contributed their time and effort to make DAS 2022 a first-rate event for the community.

In addition to the contributed papers, the program also includes two invited keynote presentations by distinguished members of the research community: Andreas Dengel from the German Research Center for Artificial Intelligence (DFKI, Germany) and Adam Jatowt from the University of Innsbruck (Austria).

We furthermore would like to express our sincere thanks to the tutorial organizer, Himanshu Sharad Bhatt from American Express AI Labs, for sharing his valuable scientific and technological insights. Special thanks are also due to our sponsors IAPR, the L3i Laboratory, AriadNext, Esker, IMDS, GoodNotes, Yooz, MyScript, ITESOFT, TEKLIA, VIALINK, and the Région Nouvelle Aquitaine and Communauté d'Agglomération de La Rochelle, whose support, especially during challenging times, was integral to the success of DAS 2022.

The workshop program represented the efforts of many people. We want to express our gratitude, especially to the members of the Program Committee for their hard work in reviewing submissions. The publicity chairs, Richard Zanibi (USA) and Joseph Chazalon (France), helped us in many ways, for which we are grateful. We also thank the discussion group chairs, Michael Blumenstein (Australia) and Umapada Pal (India), for organizing the discussion groups, and the tutorial chairs, Rafael Dueire Lins (Brazil) and Alicia Fornes (Spain), for organizing the tutorial. A special thank you goes to the publication chair, Cheng-Lin Liu (China), who was responsible for the proceedings at hand. We are also grateful to the local organizing committee who made great efforts in arranging the program, maintaining the web page, and setting up the meeting platform with support for remote attendance. The workshop would not have happened without the great support from the hosting organization, La Rochelle University.

Finally, the workshop would have not been possible without the excellent papers contributed by authors. We thank all the authors for their contributions and their participation in DAS 2022! We hope that this program will further stimulate research and provide practitioners with better techniques, algorithms, and tools. We feel honored and privileged to share the best recent developments in the field of document analysis systems with you in these proceedings.

April 2022

Seiichi Uchida
Elisa Barney Smith
Véronique Eglin

PROGRAM COMMITTEE

General Chair
Jean-Marc Ogier, La Rochelle, France

Conference Chair
Jean-Christophe Burie, La Rochelle, France
Mickaël Coustaty, La Rochelle, France
Antoine Doucet, La Rochelle, France

Conference Committee
Program Chairs
Seiichi Uchida, Fukuoka, Japan
Elisa Barney-Smith, Boise, USA
Véronique Eglin, Lyon, France

Industrial Chairs
Vincent Poulain d'Andecy, Aimargues, France
Robin Mélinand, Nantes, France

Tutorial Chairs
Rafael Dueire Lins, Recife – Pernambuco, Brazil
Alicia Fornes, Barcelona, Spain

Discussion Group Chairs
Michael Blumenstein, Sydney, Australia
Umapada Pal, Kolkata, India

Publication Chair
Cheng-Lin Liu, Beijing, China

Publicity Chairs
Richard Zanibi, New York, USA
Joseph Chazalon, Le Kremlin-Bicêtre, France

ORGANIZING LOCAL COMMITTEE

Beatriz Martínez Tornés
Carlos Gonzalez
Damien Mondou
Dominique Limousin
Emanuela Boros
Guillaume Bernard
Ibrahim Souleiman Mahamoud
Kais Rouis
Lady Viviana Beltran Beltran
Latifa Bouchekif
Marina Dehez–Clementi
Mélanie Malinaud
Muhammad Muzzamil Luqman
Musab Al-Ghadi
Nathalie Renaudin-Blanchard
Nicholas Journet
Nicolas Sidère
Souhail Bakkali
Théo Taburet
Zuheng Ming

TABLE OF CONTENT

Text Classification Models for Form Entity Linking*

María Villota[1 [0000−0002−1457−4270]], César Domínguez[1 [0000−0002−2081−7523]], Jónathan Heras[1 [0000−0003−4775−1306]], Eloy Mata[1 [0000−0003−0538−4579]], and Vico Pascual[1 [0000−0003−3576−0889]]

Department of Mathematics and Computer Science, University of La Rioja, Spain
{maria.villota, cesar.dominguez, jonathan.heras, eloy.mata, vico.pascual}@unirioja.es

Abstract. Forms are a widespread type of template-based document used in a great variety of fields. The automatic extraction of the information included in these documents is greatly demanded due to the increasing volume of forms that are generated in a daily basis. However, this is not a straightforward task when working with scanned forms because of the great diversity of templates with different location of form entities, and the quality of the scanned documents. In this context, there is a feature that is shared by all forms: they contain a collection of interlinked entities built as key-value (or label-value) pairs, together with other entities such as headers or images. In this work, we have tackled the problem of entity linking in forms by combining image processing techniques and a text classification model based on the BERT architecture. This approach achieves state-of-the-art results with a F1-score of 0.80 on the FUNSD dataset, a 5% improvement regarding the best previous method.

Keywords: Entity Linking· Text Classification · Deep learning

1 Introduction

Forms are template-based documents that contain a collection of interlinked entities built as key-value (also known as label-value or question-answer) pairs [10], together with other entities such as headers or images. These documents are used as a convenient way to collect and communicate data in lots of fields, including administration, medicine, finance, or insurance. In these contexts, there is an enormous demand in digitising forms and extracting the data included in them [10]; the latter is a task known as form understanding [7]. The form understanding task is especially challenging when working with scanned documents due to the diversity of templates, structures, layouts, and formats that can greatly vary among forms; the different quality of the scanned document images; and, the scarcity of publicly annotated datasets [7].

* This work was partially supported by Grant RTC-2017-6640-7; and by MCIN/AEI/10.13039/501100011033, under Grant PID2020-115225RB-I00.

Form understanding consists of two steps: *form entity recognition* and *form entity linking* [7]. In the former, the spatial layout and written information of forms are analysed to localise the position of form entities and to identify them as questions (keys), answers (values), or other entities present in the form. In the latter step, the extracted entities are interlinked to understand their relationships. Several approaches have been published in the literature in order to solve both tasks. Usually, they try to take advantage of both semantic text features and layout information of the forms by combining different methods [1–3,6,8,10]. In this work, we have focused on the problem of entity linking in forms using a new method that combines computer vision and natural language processing techniques. Namely, we have proposed a new method for the task of entity linking in forms that combines image processing techniques and a text classification model based on a transformer architecture. For the text classification model, we have tested different architectures using transfer learning. The best model was obtained using the BERT architecture [4], which achieved a F1-score of 0.80 on the FUNSD dataset [7]; a 5% improvement regarding the best previous method. Finally, we have publicly released all the code and models developed in this work `https://github.com/mavillot/FUNSD-Entity-Linking`.

2 Methods

A summary of our method for form entity linking is provided in Figure 1. For each answer that is found on a given form, we identify a set of candidate questions based on their distance to the answer; and, subsequently, we concatenate the text of each candidate question with the text of the answer, and use a text classification model to determine if that combination of question and answer makes sense. Finally, if multiple questions are valid for the given answer, we take the one that is closer to the answer. For our text classification models, we have fine-tuned several transformer-based language architectures [9]; namely, BERT, DistilBert, Roberta, DistilRoberta, and LayoutLM. For fine-tuning the models, we replaced the head of each language model (that is, the last layer of the model), with a new head adapted to the binary classification task. Then, we trained the models for 6 epochs on the FUNSD dataset [7]. All the networks used in our experiments were implemented in Pytorch, and have been trained thanks to the functionality of the libraries Hugging Face, FastAI and Blur using the GPUs provided by the Google Colab environment.

3 Results

In this section, we analyse the results achieved with our method. We start by exploring the performance of the studied text classification model, see Table 1. The best model for all the evaluated metrics is obtained using the BERT architecture. This model clearly overcomes the rest by a large margin, it achieves a F1-score of 0.80; whereas, the rest of the models obtain values lower than 0.70. We additionally compare our proposed method with the existing algorithms

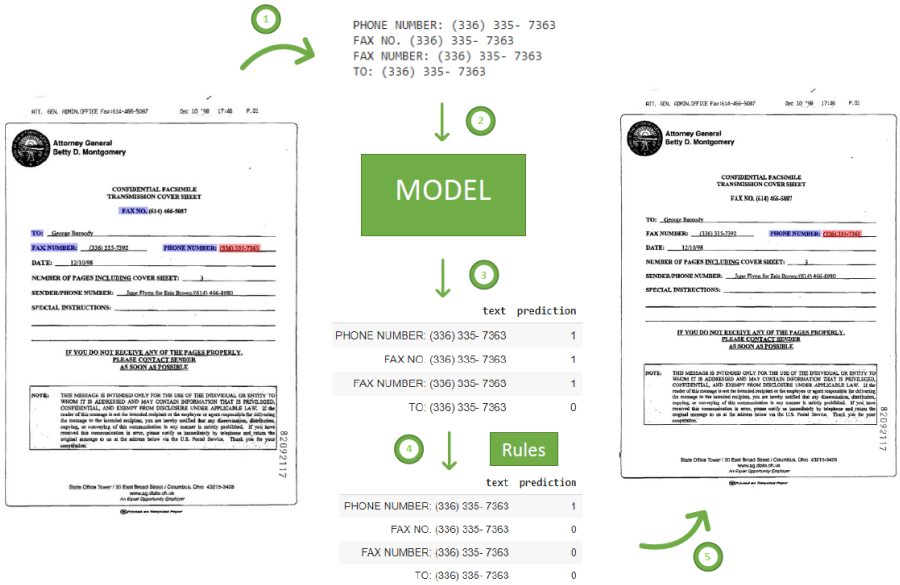

Fig. 1. Pipeline of the proposed method. (1) From an answer, a set of candidate questions are identified. (2) Each combination of candidate question-answer is fed to a text classification model that (3) identifies the valid combinations of question-answer. (4) If more than one combination is valid, the closest question is taken. (5) Finally, the results are returned.

available in the literature, see Table 1. From such a comparison, we find that the performance of our method using the BERT model improves all the existing approaches. In addition, we can notice that our method, independently of the employed text classification model, obtains a better mAP and mRank than the algorithms available in the literature. This proves the effectiveness of combining image processing techniques and deep learning models in this context.

4 Conclusion and Further work

In this paper, we have proposed a method for form entity linking based on the combination of image processing techniques and text classification models. This approach has achieved state-of-the-art results for form entity linking in the FUNSD dataset, and shows the benefits of combining deep learning models with algorithms based on the existing knowledge about documents when working in contexts where annotated data is scarce. As further work, we are interested in applying our method to more recent documents since the FUNSD dataset is formed by old documents, and also adapting our approach to work with documents written on different languages.

	mAP	mRank	F1-score
BROS [5]	-	-	0.67
Carbonell et al. [1]	-	-	0.39
FUDGE [3]	-	-	0.62
FUNSD paper [7]	0.23	11.68	0.04
DocStruct Model [10]	0.72	2.89	-
LayoutLM Word Level [8]	0.47	7.11	-
MSAU-PAF [2]	-	-	0.75
MTL-FoUn [8]	0.71	1.32	0.65
Sequential Model [8]	0.65	1.45	0.61
SPADE [6]	-	-	0.41
Ours-BERT	**0.87**	**0.49**	**0.80**
Ours-DistilBERT	0.79	0.79	0.68
Ours-DistilRoBerta	0.76	0.95	0.65
Ours-LayoutLM	0.79	0.81	0.69
Ours-RoBerta	0.77	0.94	0.66

Table 1. Comparison of our approach with existing methods for entity linking. In bold face the best results.

References

1. Carbonell, M., et al.: Named entity recognition and relation extraction with graph neural networks in semi structured documents. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 9622–9627 (2021)
2. Dang, T.A.N., et al.: End-to-end hierarchical relation extraction for generic form understanding. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 5238–5245. IEEE (2021)
3. Davis, B., et al.: Visual FUDGE: Form understanding via dynamic graph editing. In: Document Analysis and Recognition (ICDAR 2021). pp. 416–431. Cham (2021)
4. Devlin, J., et al.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics. pp. 4171–4186 (2019)
5. Hong, T., et al.: BROS: A pre-trained language model for understanding texts in document (2021), https://openreview.net/forum?id=punMXQEsPr0
6. Hwang, W., et al.: Spatial dependency parsing for semi-structured document information extraction. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. pp. 330–343 (2021)
7. Jaume, G., Ekenel, H.K., Thiran, J.P.: FUNSD: A dataset for form understanding in noisy scanned documents. In: Proceedings of the 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). vol. 2 (2019)
8. Prabhu, N., et al.: MTL-FoUn: A multi-task learning approach to form understanding. In: Document Analysis and Recognition – ICDAR 2021 Workshops. pp. 377–388 (2021)
9. Razavian, A.S., et al.: CNN features off-the-shelf: An astounding baseline for recognition. In: CVPRW'14. pp. 512–519 (2014)
10. Wang, Z., et al.: Docstruct: A multimodal method to extract hierarchy structure in document for general form understanding. In: Proceedings of the Findings of the Association for Computational Linguistics (EMNLP 2020), pp. 898–908 (2020)