Pattern Recognition Letters 166 (2023) 76–79

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Reflections of an ancient document processor

George Nagy*

Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, United States

ARTICLE INFO

Article history:
Received 22 July 2022
Revised 3 November 2022
Accepted 9 January 2023
Available online 10 January 2023

Edited by: Maria De Marsico

ABSTRACT

The bulk of the documents that affect our lives are digital or born digital. Our laborious investigations of layout, script, font and graphics, are turning into mere exercises with little influence on pursuits outside the Document Analysis and Recognition (DAR) community. Recent performance improvements on such tasks, even if based on deep learning and AI, are as much the result of advances in computer hardware as of breakthroughs in document research. It is time to automate tasks beyond transcription. This Commentary addresses our mission, our approach to some technical issues, and the role of AI in DAR. Opportunities for a wider role for document analysis include more pervasive application of statistical decision theory, integrated genre analysis, summarization, interpretation and information extraction, bolder goals in content analysis, and alternative modalities, induced by the open source movement, for sharing research results. Importantly, expanding the scope of our research incurs increased responsibility for retaining human prerogatives in critical decision making and preserving essential human skills like good writing and discriminative reading.

1. Alert

Given a few factoids – true or false – in the form of tables or lists, an AI[1] text generator can convert them into grammatical, and even stylish, prose. Given the narrative, another AI app can extract the factoids and stuff them into a database. The cycle of structured-data-to-narrative-to-structured-data can be extended ad libitum. Computers writing for computers! According to recent evidence, AI may be more reliable, and certainly faster than humans, at separating verified facts from rumors, conjectures and outright falsehoods. These and other advances in reading and writing tasks fairly compel the Document Analysis and Research community (and our IAPR relatives) to expand our technical horizon and also stake out an independent and indispensable role for humans.

2. Note

This is a commentary, not a technical or scientific report, a survey, or a review. Instead of detailed references, it has only a bibliography of about fifty items that prompted or shaped these reflections. Many entries point not to technical articles but to newspaper or magazine articles or web postings that may reflect immature and evolving ideas.

3. Documents

While there is a long history of changing perspectives on what a document *is* or *is not*, the Wikipedia's broad (and lengthy) definition is well aligned with our proclivities. It includes written, drawn, presented, or memorialized representations of fictional and nonfictional thought, as well as their electronic embodiment in computer files. An older and more abstract definition, *any concrete or symbolic indication, preserved or recorded, for reconstructing or for proving a phenomenon, whether physical or mental*, would also serve. Most current applications of document processing can be found on the ICDAR, DAS and IJDAR web sites, and on the CFP for this special issue.

Types of documents unfamiliar to 19th C librarians include email, short text messages, electronic journals, eBooks, blogs, podcasts, videos, visualizations, and computer art. The rapidly shifting nature and role of documents suggests a complementary shift in research paradigms. Instead of categorizing documents by appearance and encoding (i.e., by language, glyphs and formating), which are all mutable in the digital world, we should consider document *content* characteristics that are more directly related to higher-level human endeavors like planning, decision making and knowledge

* Corresponding author.
E-mail address: nagy@ecse.rpi.edu

[1] "AI" is used here in its current sense of a computer program that combines sorting, arithmetic and logic operators in a configuration (usually a neural network) that yields an unpredictable but mostly useful outcome. In the early 1960's, the term *artificial intelligence* was often used for automata and rule based approaches to problem solving in contrast to *connectionism* (neural networks).

https://doi.org/10.1016/j.patrec.2023.01.006
0167-8655/

discovery. None of the corresponding document processing objectives listed below is entirely new, but they have hitherto seldom surfaced in DAR.

4. A brief history

The first successful application of document processing was Optical Mark Recognition (OMR). Automated *symbol* recognition did not surpass human ability until the advent in the 1950′s of the durable MICR E13B font for bank checks. Electronics, and then computers, progressed to OCR-A, OCR-B, monospaced and proportional typewriter fonts, typeset text, incunabula, illuminated manuscripts, hand-printing, handwriting, signatures, logos, tables, graphics, comics, and scene-text, in Latin and other scripts.

Postal recognition evolved from ZIP codes to complete address reading. Among the earliest targets of information extraction from semi-structured text were factoids from obituaries and used-car ads. Narrative summarization hit a snag because the available ground truth was so poor. There were few attempts at *goal-oriented* summarization of free-flowing text. Human summaries often consisted only of a few sentences lifted from the Introduction and the Conclusion. Current authors (or their software?) still resort to this inappropriate expedient in preparing their Abstracts, thereby substituting motivation for substance.

The performance of today's automated story understanding software is comparable with that of top high-school students. Conversely, text generators can now turn numerical weather data, sports scores, and stock market quotations into narratives barely, if at all, distinguishable from the product of human scriveners.

With the currently available translation programs, or even with just the common spelling, grammar and style checkers, fluency is no longer necessary to write in a foreign language. Even technical knowledge is becoming dispensable for aspiring authors. With large public repositories of test data and toolboxes offering an infinite variety of algorithms, any computer-literate aspirant to authorship can submit a plausible draft to a journal or conference.

5. Provenance

Authorship analysis has always been part of literature. On the web, documents often lose their integrity and their links to their authors. If the disconnect affects their credibility, then these document fragments must be traced to their source. Already available tools include signature and scribe recognition, encryption, forensic typeface analysis, digital signatures and watermarking, and plagiarism detection. *Adversarial learning* improved spam filters. We don't yet have measures of *conceptual similarity* that can distinguish shared origins from similar conclusions reached by independent workers.

The effects of false provenance induced by malice range from annoyance, misinformation and loss of privacy to outright fraud and catastrophic malfunction. Document provenance monitoring can be improved by including all available appearance, semantic content, metadata and routing tags.

Can we still tell AI-generated text from human utterance? We could build classifiers that do this, and even keep up with changes in both human and machine writing as new samples of both are posted on the web. What about machine translation from one language to another vs. human translation? Computer vs. human proofs of mathematical theorems and algorithm verification? Medical and nutritional advice? Instead of passing the buck, perhaps we can join the data-integrity efforts of Big Data consortia in the health sciences, meteorology and astronomy.

6. Genre recognition

Every document is created with some purpose that determines its *genre*. An automated cataloguer should be able to differentiate invoice from receipt, questionnaire from legal notice, biography from novel, news from propaganda, science from science fiction, prognosis from diagnosis, poetry from doggerel, or even erudition from parody. Perhaps AI also needs to be endowed with a sense of humor. Even more problematic is AI's sense of *time*, including urgency (for author or reader) in the life of a document.

It may also be useful to recognize documents and messages meant to influence, rather than to inform, entertain or annoy. Current efforts on *emotion assessment* barely scratch the surface. Finer classification would reveal whether the sought behavior is social, political, or financial for both the source and the recipient. It is understood that most such distinctions are ambiguous and intrinsically multi-dimensional. Fortunately we have ample experience with multi-valued, multi-dimensional features.

7. Document synthesis

PC software has always generated hints, suggestions and error messages. These evolved into spelling, grammar and style checkers, next-word prediction, educational videos, infomercials and advertorials. They may be mediated by digital assistants like Siri, Cortana, Google Assistant or IBM Watson, that (who?) can be difficult to differentiate on chat platforms from pre-programmed humans. Messages are customized not only according to one's current activity, but also yesterday's or last month's. How can we apply the benefits of advanced techniques of document synthesis to improve the human condition?

8. Integrated document processing

Decision-support systems have traditionally obtained all the information necessary to make a good call from a user-designated database, collection of semantic triples, or ontology. A plethora of useful facts, plans, and calculations may, however, be scattered in other accessible venues. Automated systems must therefore learn, as we do, to assemble their own reading lists.

Document processing programs must also learn on what (not only whom!) to call for help when stuck, to *network* with other document processing software, and to collaborate rather than compete. Shades of classifier combination and of transfer-learning! Immature apps should be able to get help from computer tutors with no functionality besides serving as information conduits. Currently only the largest organizations can afford to develop independent yet collaborative systems, but IAPR should think about it too. Should we expect more from AI than from our fellow humans? Reliable automated content analysis should not embarrass us any more than accurate computer arithmetic. AI, after all, has speed and memory on its side.

Can we develop a *theory* linking various types of computer-born documents to their content instead of only conducting experiments that improve some performance measure on an arbitrary selection of traditional documents? We could make better use of ancillary data and large-scale context instead of re-using established methods by converting computer-native images to pixel-arrays. Scene-text reading might benefit from GPS data, from exposure and other settings preserved with the image, from analyzing other photos on the same camera-roll, and perhaps even from the web profile of the clicker. Medical image analysis should also draw on properties of other documents in the same collection (articles from a particular journal, cell images from the same laboratory run, and CAT scans of patients with similar conditions).

Open Source Software has given us Linux, LaTeX, python, Octave, (Apache) OpenOffice, LibreOffice, and most C compilers. Peer review of proposed improvements in OSS seems to be faster and more open than in our publications. The product of these endeavors is a comprehensive whole rather than a fractional contribution. Some larger DAR goals would benefit from cohesive, large-scale, long-term collaborations like the above.

9. Cost-driven classification

To remain relevant, DAR cannot stop at word error rates. The DAR community has been too slow to explore broader relationships between humans and documents. Most published classification results are still based on error and reject rates, sensitivity and specificity, precision and recall, the F1 score, Type I and II errors, and the ROC curve, which were all designed for two-class problems. Modifying them for multiclass problems with highly unbalanced memberships and costs tends to be clumsy and ineffective.

Statistical Decision Theory (developed c. 1950) lets us consider financial and social costs. It advocates minimizing risk (the expected loss) as the ultimate objective. This requires sound sampling and class-specific cost functions. Are web tools already available to periodically construct a plausible *document census* from a DAR perspective? Can search engines be modified to provide unbiased samples of specific, countable document populations? Can we collaborate with economists and with medical administrators, museum curators, publishers and others with budgetary responsibilities, to begin to fill the cost function lacunae in document analysis?

10. Reproducibility, generalizability and self-learning

Some definitions of *reproducibility* imply reproducing conclusions of earlier work on new data by a different team. What is the connection with the long-established pattern recognition and machine learning notions of *generalizability, adaptation,* and *unsupervised learning*?

11. Dissemination of research results

Research results on topics closely related to document analysis are scattered among journals and proceedings in the Natural Language Understanding, Linguistics, Information Retrieval, Knowledge Extraction, Data Mining, Information Retrieval, Pattern Recognition, Machine Learning and Neural Network houses of AI. Recent issues prompted the emergence of Explainable AI (XAI), Fair, Accountable, Confidential and Transparent algorithms (FACT) and Replicable and Reproducible Pattern Recognition (RRPR). Balkanization appears to be more prevalent in research on artifacts than in the natural sciences.

Journal readership has been steadily declining in spite of the publishers' efforts to bring their offerings to our attention based on our web profile and activity. The median number of readers of a technical article is in the single digits, and that of conference papers is even less. How could it be otherwise with 1.5 million peer-reviewed articles published annually in some 30,000 academic journals? Elsevier and Springer each publish more than 2500 journals. Only a minority of the published articles are ever cited. It is estimated (I don't know how) that about 20% of the articles cited have actually been read by the citing author.

Like other vendors, publishers implore our feedback to improve their marketing. Professional societies send us questionnaires whenever we participate in their activities. Journals are adding graphical abstracts, research highlights, video and audio to make our communications more distinguishable. Premature conference submission deadlines with onerous formating requirements are postponed repeatedly. Acceptance decisions are delayed. Proceedings, that used to be available before the opening session, are months late. Such practices do not reflect well on our foresight and reliability. We cannot blame *everything* on COVID!

Appendices of journal articles, formerly a few pages of supplementary information, are turning into vast on-line repositories of computer code and data that are virtually impenetrable to inspection. The complexity of off-the-shelf AI software prevents researchers from comprehending and presenting potentially important aspects of their experiments. In the MATLAB Deep Learning toolbox, for example, the critical parameters of the algorithms are shrouded by layers of nested function calls. The code in the P-files of the innermost layers, where most of the action takes place, is invisible to mere license holders.[2]

Although there are plenty of statistics about social network demographics and usage, it is difficult to find any estimates of the fraction of messages actually seen by human eyes. Never having joined a social network, I am bereft of ideas for research on these vast channels and receptacles of contemporary thought.

Charges for open source publication and rapidly growing conference registration fees are a major source of income for professional societies, publishers and institutes. The costs are often passed on to universities, research laboratories, and granting agencies. In 2018, the international market for scientific publications accounted for about nine billion dollars per year. Yet most of the work (writing, peer review, editorial decisions, proofreading and editing) is carried out by the researchers themselves, and the cost of storing and disseminating electronic documents has never been lower.

Having spent a lifetime in research and related reading and writing, I am dismayed about this sorry state of affairs. There must be more effective ways of consolidating and disseminating research results than thorough our run-away journal and conference proceedings! I have no real solution to offer, and can only draw attention to a few plausible directions for DAR.

We can encourage posting preprints on public depositories like arXiv and Zenodo. Most journals allow this practice. Some communities have established volunteer and completely open peer review systems (e.g. *Peer Community In*) for publicly accessible reports. DAR is still small enough to allow launching such a project quickly.

Given the reach of web search engines, we could dispense with almost identical sections on prior work in experimental reports on the same data. We could develop and deploy content analysis to verify the relevance of each reference. This could be added to the automated format checking and plagiarism detection software. We should commission surveys that adhere to the rigorous standards of *systematic review* and *meta-analysis* in the health sciences. Yet we cannot even agree on family-name first or last in citation formats!

Digital libraries and institutional repositories rendered our work more *accessible*, Linked Open Data (LOD), Uniform Resource Indicators (URI), the Resource Description Framework (RDF), machine readable metadata, ontologies, and knowledge graphs –all under the aegis of the Semantic Web and W3C – will make research results more *discoverable*. Is DAR pulling its weight to make all this work? Technical communication problems are not unique to our community, but we bear special responsibility because *documents* have always been the vehicle for sharing knowledge.

Many industrial research labs have revised their reward structures to reduce gratuitous publication, but Academia just keeps counting. Well-established senior members of our community should exert their individual and collective influence to lessen

[2] P-code files are an obfuscated, execute-only form of MATLAB code.

the influence of the h-index and impact-factor games in evaluating research accomplishment. (The most reliable alternative that I have seen are detailed and explicit letters of recommendation from disinterested parties.) Conference travel reimbursement should be made as available to those who listen as to those who talk. I fear, however, that senior researchers who made their way under the existing system are likely to take a conservative stance against major change.

12. Assigning responsibility

One person's recyclable may be another's archival treasure. Personalized document processing is impossible without knowing the characteristics of the recipients or potential recipients. Such characteristics may include an individual's circumstances, disposition, education, experience, and goals. If automated analysis is to help decide what to read, study, share, and discard, then document features must be combined with reader features other than propensity to buy, rent, lease or pay bills on time. However, neither questionnaires nor covert surveillance are appealing ways of collecting personal information. Under what conditions is it appropriate to entrust document processing systems with our intellectual attributes?

The distinction between AI and human thought is worth preserving. Responsibility for judging humans (in education, employment, judicial jurisdiction) should never be delegated to computers. Nor should judging values in fine arts, music or literature. The DAR community cannot neglect the human impact of its work. We must find out how computers can help us without absolving us from responsibility for their conduct. Document processing deserves no less serious attention to ethics than drug discovery or programming robots that manipulate tools and weapons.

13. Envoy

Fifty years ago I kept hearing and reading that OCR was a solved problem. There may be those who believe now that DAR is passé (though the US Chamber of Commerce estimates that the Government spends $39 billion on processing paper forms). Who are we and where are we going? Although I am uneasy with some of the current developments, I am confident that the next generation will not suffer from a dearth of exciting quandaries. Documents may evolve, as they have for millennia, but they are here to stay.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

This is a short Commentary. No data is involved.

Acknowledgment

The cogent suggestions of the two PRL reviewers contributed significantly to this Commentary.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.patrec.2023.01.006.

Bibliography on next page.

Bibliography

What is a document?

D. Doermann, E. Rivlin, A. Rosenfeld, The function of documents, Image and Vision Computing 19, 11, 789-814, August 1998.

S. Ferrilli, Automatic Digital Document Processing and Management, Springer, 2011.

Mi. Buckland, What Is a "Digital Document"? In Document Numérique, Paris. 2(2). 221-230. 1998. http://people.ischool.berkeley.edu/~buckland/digdoc.html accessed on October 9, 2022

Wikipedia: Document https://en.wikipedia.org/wiki/Document accessed on October 9, 2022

Optical Character Recognition and Digital Image Analysis

H.F. Schantz, A History of OCR: Optical Character Recognition, Recognition Technologies Users Association, 1982.

S.N. Srihari,Y-C Shin,V. Ramanaaprasad, Z. Shi, Document image-processing system for name and address recognition, Int'l J. of Imaging Systems and Technology, 7, 4:. 379-391, 1996.

H. Bunke and P.S.P. Wang, Character Recognition and Document Image Analysis, World Scientific 1997.

I. H. Witten, A.r Moffat, & T.y C. Bell, Managing Gigabytes, Morgan Kaufman, 1999.

G. Nagy, Twenty Years of Document Image Analysis in PAMI, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, #1, (invited) 20th Anniversary Issue, pp. 38-62, 2000.

M. Cheriet, N. Kharma, C_L Liu, C.Y. Suen, Character Recognition Systems, Wiley, 2007.

S. Marinai and H. Fujisawa (Editors), Machine Learning in Document Analysis and Recognition, Springer 2008.-

D. Doermann and K.l Tombre, Editors, Handbook of Document Image Processing and Recognition, Springer-Verlag London 2014, 2019.

Document Aanalysis dnd Recognition

Li Xu and David W. Embley, Categorisation of web documents using extraction Ontologies, Int. J. Metadata, Semantics and Ontologies, Vol. 38, No. 1, pp. 3-19, 2008.

D. Impedovo and, G. Pirlo, Automatic signature verification: The state of the art, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 609-635. 2008.

S. Eskenazi, P. Gomez-Krämer, J_M. Ogier, JM. When Document Security Brings New Challenges to Document Analysis. In: Garain, U., Shafait, F. (eds) Computational Forensics. IWCF IWCF 2012 2014. Lecture Notes in Computer Science, vol 8915. Springer, 104-116, 2015.

G. Nagy, Disruptive developments in document recognition, Pattern Recognition, Vol. 79, 106-112, August 2016.

O. Tas, and F. Kiyani, A Survey of Automatic Text Summarization, 2nd World Conference on Technology, Innovation and Entrepreneurship, May 12- 14, 2017, Istanbul, Turkey. Edited by Sefer Şener, PressAcademia Procedia, pp.204- 213, 2017.

K Ubul, G Tursun, A Aysa, D Impedovo, G Pirlo, T Yibulayin, Script identification of multi-script documents: a survey, IEEE access 5, 6546-6559, 2017.

R. Plamondon, A. Marcelli, M.A. Ferrer, The Lognormality Principle and its Applications in E-security, E-learning, and E-health, World Scientific Machine Perception and Artificial Intelligence Vol 99, 2021.

M. Gambhir, V. Gupta, Recent automatic text summarization techniques: a survey. Artif Intell Rev 47, 1–66 (2017). https://doi.org/10.1007/s10462-016-9475-9 . accessed on October 9, 2022

J. R. Morrey, Extracting Information from Historical Genealogical Documents, Towards Data Science, January 19, 2022, https://towardsdatascience.com/extracting-information-from-historical-genealogical-documents-ab3068b10715 accessed on Oct. 9, 2022

Document synthesis

L. Dou, G. Qin, J. Wang, Jin-G. Yao, and C-Y Lin. 2018. Data2Text Studio: Automated Text Generation from Structured Data. In Procs. 2018 Conf. on Empirical Methods in NLP System Demonstrations, 13–18, Brussels, Belgium. Ass'n for Computational Linguistics, 2018,

S. Johnson, The Writing on the Wall, The New York Times Magazine April 17, 2022.

IBM Watson Patient Synopsis, https://www.ibm.com/products/watson-imaging-patient-synopsis ,accessed on October 9, 2022

MarketWatch Automation, https://www.marketwatch.com/author/marketwatch-automation, accessed on October 9, 2022

United Robots: High end automated content in any team sport, https://www.unitedrobots.ai/content-services/sports accessed on Oct. 9, 2022

C. Bean, Automated Sports News: The Good, the Bad and the Unlikely, Medium, Dec. 30, 2018. https://medium.com/@ConnorBean6/automated-sports-news-the-good-the-bad-and-the-unlikely-533115efd319 accessed on Oct. 9, 2022

Neural Networks

E. R. Caianiello, Outline of a theory of thought-processes and thinking machines, Journal of Theoretical Biology 1 (2): 204–235, 1961.

F. Rosenblatt, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Spartan Books, 1962.

Y. Bengio, A. Courville and P. Vincent, Representation Learning: A Review and New Perspectives, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, Aug. 2013.

I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2015.

Y. LeCun, J. Bengio and G. Hinton, Deep learning, Nature, Vol 521 pp. 436-442, May 28, 2015.

G. Lewis-Kraus, Going Neural, The New York Times Magazine, Dec. 18, 2016,

C. Metz, Quietly hoarding millions of faces culled from the web, New York Times. 7/14/2019.

Artificial Intelligence

M. Minsky,. Steps toward artificial intelligence. Proceedings of the IRE 49.1 (1961): 8-30.

A.. Gamba, New Developments in Artificial Intelligence and Pattern Recognition, in Computer and Information Sciences-I (J. T. Tou and R. H. Wilcox, eds.), pp. 219–229, Spartan Books, Washington, D. C., 1964.

R.E. Neapolitan, Probabilistic Reasoning in Expert Systems, Wiley, 1990.

R. Kurzweil, The Singularity Is Near: When Humans Transcend Biology, Penguin Random House, 2006.

N. Nilsson, The quest for artificial intelligence: A history of ideas and achievements, Cambridge University Press, 2010.

C. Kuang, Can A.I. be taught to explain itself? The New York Times Magazine, November 26, 2017.

B. M. Lake, R. R. Salakhutdinov, J.Tenenbaum, One-shot learning by inverting a compositional causal process, Neural Information Processing Systems 26, 2013.

T. Simonite, Teaching Machines to Understand Us, MIT Technology Review, Aug. 6, 2015.

Eliza Strickland, The Turbulent pas and uncertain future of AI, IEEE Spectrum, October 2021. The New York Times, July 14, 2019

R. Kurzweil, Ray Kurzweil on How We'll End up Merging with Our Technology, The New York Times Book Review, March 19, 2017,

M. Hudson, Taught to the Test, Science, Vol. 376, Issue 6593, May 6, 2022.

S. Russell & P. Norwig, Artificial Intelligence: A Modern Approach, 4th Edition, Pearson 2021,

C. Metz. A.I. Does not have Thoughts, No Matter What You Think, The New York Times, AugKust 7, 2022.

Publishing modalities

K. K. Landes, A Scrutiny of the Abstract, Bulletin of the American Association of Petroleum Geologists, Vol.35, #7, pp. 1660-1660, 1951.

J. Hendler, Reinventing Academic Publishing-Part 1, in IEEE Intelligent Systems, vol. 22, 5, pp. 2-3, 2007.

J. Hendler, Reinventing Academic Publishing, Part 2, in IEEE Intelligent Systems, vol. 22, 6, pp. 2-3, 2007.

J. Hendler, Reinventing Academic Publishing, Part 3 in IEEE Intelligent Systems, vol. 23, 1, pp. 2-3, 2008.

A. K. Biswas And J. Kirchherr, Prof, no one is reading you, The Straight Times, April 11, 2015.

International Association of Scientific, Technical and Medical Publishers, The STM Report, An overview of scientific and scholarly journal publishing, Fourth Edition, 2015. https://www.zbw-mediatalk.eu/wp-content/uploads/2017/07/STM-Report.pdf accessed on Oct.9, 2022.

P. G. Altbach and Hans de Wit, Too much academic research is being published, University World News, Sept. 7, 2018, https://www.universityworldnews.com/post.php?story=20180905095203579 accessed on October 9, 2022.

T. Guillemaud, Benoit Facon, Denis Bourguet. Peer Community In: A free process for the recommendation of unpublished scientific papers based on peer review. ELPUB 2019 23rd edition of the International Conference on Electronic Publishing, Marseille, France. Jun 2019.

A. Grudniewicz, D. Moher, K.y D. Cobey and 32 co-authors, Predatory journals: no definition, no defence, Nature | Vol 576 | 12 Dec.2019

Adam Ruben, Why scientific journal authorship practices make no sense et al., Science, Oct. 28, 2021,

Reproducible research

R. D. Peng, Reproducible Research in Computational Science, Science 334, Issue 6060, pp. 1226-1227, Dec. 2, 2011. Available at NLM PubMed Central 22144613, Dec. 2, 2011.

M. Colon, B. Kerautret, A. Krähenbühl, An Overview of Platforms for Reproducible Research and Augmented Publications, RRPR 2018, pp. 25-49, June 2019, https://www.researchgate.net/publication/334623730_An_Overview_of_Platforms_for_Reproducible_Research_and_Augmented_Publications accessed on October 9, 2022

B. Kerautret, M. Colom, A. Krähenbühl, Daniel Lopresti, Pascal Monasse, H. Talbot: Reproducible Research in Pattern Recognition - Third International Workshop, RRPR 2021, Virtual Event, January 11, 2021, Revised Selected Papers. Lecture Notes in Computer Science 12636, Springer, 2021.