A new type of writer

The first book by AI author »B. Writer«

By Jan Schwenkenbecher

Together with his team, Frankfurt computer linguist Christian Chiarcos has developed a software. This software has written the first machine-generated science book.

At first glance, the book »Lithium-Ion Batteries« seems to be a completely normal book: It was penned by »Writer, B.« and appeared in the spring of 2019. It was published by scientific publishers Springer Nature. It has a title, a subtitle, four chapters and a large number of subchapters. There is a preface, a bibliography and the fact that the headings sound a little awkward does not necessarily confuse the reader – after all, it's a textbook. Springer Nature found the book's publication worthy of a press release nonetheless, and this has less to do with the content than with »Writer, B.«, the author.

The »B« stands for Beta and Beta Writer is not the name of a battery researcher given that name at birth by slightly eccentric parents. Beta Writer is not a chemist nor a researcher, Beta Writer is not even a human being. The explanation can be found on page 4 of the book: »This book was machine-generated.« The book »Lithium-Ion Batteries« is the first book composed by artificial intelligence that Springer Nature has published. Beta Writer is the software that wrote it. And in this sense Beta Writer also has parents: A research team led by Christian Chiarcos, professor for applied computer linguistics and head of the Applied Computer Linguistics (ACoLi) working group at Goethe University,

and Niko Schenk, a doctoral researcher in his group. Chiarcos will say later of Beta Writer: »Beta Writer is the name of an algorithm that we've developed here on the basis of our own already existing work and that of the community in order to generate books. The plan was to use it to produce the first machine-generated science book. And we've done just that.«

Selecting sources

Chiarcos is also the right person to ask exactly how Beta Writer wrote the book. He can be found at the Institute of Computer Science in the Frankfurt suburb of Bockenheim – woolly jumper, jeans, spectacles, brown hair that blends into one with a full brown beard – first of all he gets himself a big cup of coffee from the machine then sits down on one of the four low armchairs. There is a shelf full of books along one wall, along another two whiteboards on which countless formulae are scribbled, a pile of documents on the desk. Chiarcos picks up a pen and draws diagrams in the air when explaining the four steps he and his colleagues took along the route to their AI book.

»The first step is pre-processing, where you start by building up a collection of possible sources,« he says, explaining the basis on which the programme composed the book's content. This can be PDFs or Word or XML documents. »We then filtered these sources according to specific keywords given to us by experts in the field in question,« says Chiarcos. »In this way we chose the scientific publications most suita-

Machines have what it takes to be book authors and can create a literature overview in science books.

Lithium-ion batteries are the subject of »B. Writer's« book.

ble for the book.« The researchers extracted the text from these documents, which was not so easy as there were all sorts of chemical formulae between the words and the punctuation marks. However, they mastered this challenge and at the end were left with a collection of 1,086 publications, all written in English and from Springer Nature's library.

»B. Writer« is the author

In the second step, the researchers used different methods to create a structure for the new book from this collection of texts: Structure generation. Computer linguist Chiarcos explains:

About Christian Chiarcos

Christian Chiarcos, born in 1977, computer scientist and linguist, earned his doctoral degree in 2010 at the University of Potsdam with a thesis on computer linguistics.
In 2012, he organized the first workshop on »Linked Data in Linguistics« in Frankfurt. He was then a visiting researcher at the Information Sciences Institute of the University of Southern California in the USA. In 2013, he was appointed as Junior Professor for Applied Computer Linguistics at the Faculty of Computer Science and Mathematics of Goethe University. In parallel to his junior professorship, since 2015 he has headed the »Linked Open Dictionaries« Early Career Research Group, which is funded by the Federal Ministry of Education and Research.

chiarcos@informatik.uni-frankfurt.de

»For all the documents, we identified their relative similarity to each other.« Whereby similarity referred to how each respective text resembled itself. »The most similar ones are clustered together until you have a tree structure.« Things with little similarity drop out, the user can specify how many chapters, sections and subsections he wants in the end and also how much text is available for Beta Writer in each subsection in order to put together the respective publication.

»The actual text generation, the third step, then comprises identifying the most important statements within a text,« explains Chiarcos. To do this, he tested various approaches with his colleagues: A classic, graph-based technique, a modern neural model – at the end they used the different methods in parallel. They tested various quantifiers in a number of runs and looked at which result the experts in the subject preferred. These authorities were chemistry and battery experts at Springer Nature. Several times over, Chiarcos and his colleagues presented different variants of interim results of what Beta Writer had so far put together to the experts. The experts rated content and style – whereby they, as can easily be seen when reading the book – placed more weight on technical accuracy than on elegant language.

Sentences are reworded

On the basis of this feedback, Chiarcos and his team then also weighted their methods quite differently depending on where they were applied in the book: The introductory texts of each chapter, which Beta Writer compiled from all the publications contained in it, have a specific weighting. The subsections, in which a single publication is summarized respectively, have a different one. The researchers weighted their methods differently once again for the chapters »Summary« and »Applied Research«.

The text is compiled as follows: »We take a complete sentence,« says Christian Chiarcos, »we perhaps eliminate parts of it, we substitute other parts, we rearrange it on the basis of syntactic analysis.« If the resulting sentence is sufficiently different from the original one, it is not marked as a citation. The authors of the original sentences do not need to worry about plagiarism anyway. Even if the new sentence is not cited verbatim in the book, the corresponding footnote with the source is always indicated after it.

Critical comments too

Finally, the last step for Chiarcos and his team was post-processing. They collated all the references in the bibliography, inserted the chemical formulae again, which had previously been sub-

stituted with wildcards in order to work on the text, converted the document into a format readable for the scientific publishers, Springer Nature, and presented it to them.

And those responsible there seem to be giving Beta Writer's debut novel the best reviews. Alongside Christian Chiarcos and his colleague Niko Schenk, Henning Schoenenberger, Director Product Data & Metadata Management at Springer Nature, wrote part of the book's introduction and was not sparing with praise and pathos: »This book on lithium-ion batteries has the potential to ring in a new era of scientific publishing,« says Schoenenberger. The future will show whether that will actually happen. Now, after about a year, the book has been cited 14 times and downloaded 357,000 times. But it is free of charge.

The download figures should not, however, belie the fact that the project was also criticized. »The feedback we received was indeed mostly positive,« says Chiarcos. But there were also some very critical comments regarding the question of social and political responsibility. »People stressed the responsibility of science and asked whether the system didn't produce a distorted picture of a subject area, a bias. «

**A finer language, more attractive headlines**
Indeed, the publications included are selected on the basis of their similarity to each other. Now if these original data already distort reality because, for example, someone has provided extensive funding for a specific research topic or a specific research group and a particularly large number of publications are now available in this subdomain – then the system reproduces this distortion and amplifies the bias. »Although our system does not produce such a distortion,« says Chiarcos, »there is no way to compensate for it automatically. This can only be done by an expert in the field in question getting down to work and reviewing the literature manually.«

Besides that, there are numerous other aspects that Christian Chiarcos would like to optimize together with his colleagues. A finer language. More attractive headlines. Greater coherence. Moreover, there are other research fields in science apart from lithium-ion batteries where Beta Writer could also compile the one or the other anthology.

The book now published has answered the question whether artificial intelligence can write scientific books. It can. The follow-up question is now: In which role will Beta Writer – or similar algorithms – find its place in the libraries of scientific publishers? Will there be a review now and again? Or will all the non-fiction authors in the country soon be out of work?

The real strength of Beta Writer does not lie in the fact that it has written a scientific book. It is the fact that it wrote a scientific book about a random research topic and the users – in this case Christian Chiarcos and his colleagues – could tell the programme how many chapters they would like and how long they should be. Perhaps Beta Writer's main work could thus become something completely different to writing books. After all, it is a software that can automatically create a highly individual literature overview. This is what researchers need, for example, when tackling a new topic, but it is also what doctoral researchers need when writing their thesis.

»That is indeed what I consider to be the most probable application for this technology in the long run,« says Christian Chiarcos too. He believes that people will not use the software as a tool for generating texts »but rather as a tool to help them write books more effectively.« ●
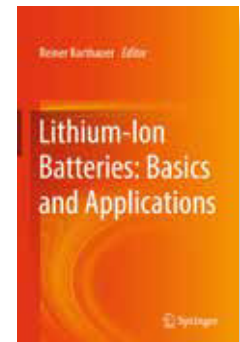
»B. Writer«
Reiner Korthauer, Hg.
**Lithium-Ion Batteries:
Basics and Applications**

The first book by AI author »B. Writer«.
Free download under
https://tinyurl.com/BWriterBattery

In response to current events, another book by »B. Writer«, a literature overview on SARS-CoV-2 and COVID-19: https://tinyurl.com/BWriterCovid

The author

Jan Schwenkenbecher is a freelance science journalist and lives in the Rhine-Main region. He studied psychology in Giessen and Mainz and then learnt the art of journalism in the framework of practical training at the Süddeutsche Zeitung newspaper.

jan.schwenkenbecher@posteo.de