NDS Writer



TECHNISCHE UNIVERSITÄT DARMSTADT

Annotation tool for creating high-quality multi-document summarization corpora



IVE PREPARATION OF INFORMATION FROM HETEROGENEOUS SOURCES

Christian M. Meyer, Darina Benikova, Margot Mieskes, and Iryna Gurevych Research Training Group AIPHES, Technische Universität Darmstadt, Germany



- Amount of textual information (e.g., in the web) becomes intractable
- We need automatic summarization!
- Single-document summarization cannot reduce the *number* of documents
- We need multi-document summarization!
- Corpora are crucial for training and ulletevaluating automatic methods
- Existing corpora limited in terms of size, domains, genres, languages
- We need summarization corpora!

- Summarization is one of the most challenging NLP tasks, as it requires solving multiple subtasks, incl. content selection, redundancy removal, coherent writing
- Existing corpora focus on a single subtask (e.g., the final summary text), which prevents us from evaluating intermediate steps
- Currently no freely available tools for modeling a complex multi-document summarization setup and storing intermediate results and system-user interactions
- We need tools for complex annotation setups!

<text></text>													
ID Topic Documents Step	MDSWriter Dashboard				Dashboard User: Demo (Logout)						MDSWriter: Albert Einstein 1. Nugget identi		
1 Atbert Einstein 10 Image: Second Sec	ID Topic	Documents	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7		Wikipedia: Albert Einstein		
 A laphaGo vs. Lee Sedol 11 A laphaGo vs. Lee Sedol 12016 A daphaGo vs. Lee Sedol 13 A daphaGo vs. Lee Sedol 13 A laphaGo vs. Lee Sedol 14 A laphaGo vs. Lee	1 Albert Einstein	10	✓	-	-	-	-	-	✓		Wikipedia: Albert Einstein Albert Einstein was a German-born		
3 Zika Virus 9 Image: Construction of the second cons	2 AlphaGo vs. Lee Sedol	11	-	-	-	\checkmark	-	-	-		theoretical physicist. He developed general theory of relativity, one of		
4 Brexit 25 Image: Comparison of the comparison of the product of the comparison of the comp	3 Zika Virus	9	\checkmark	-	-	\checkmark	\checkmark	-	0		two pillars of modern physics (along quantum mechanics). Einstein's wo		
 	4 Brexit	25	-	-	-	\checkmark	0	×	×		also known for its influence on the philosophy of science. Einstein is be		
 	5 46th World Economic Forum	15	\checkmark	-	-	0	×	×	×		known in popular culture for his mass-energy equivalence formula E		
7 Refugee Crisis 20 0 X X X 8 Rio Olympics 15 0 X X X 9 Myanmar general elections 15 0 X X X 10 Helmut Schmidt 11 0 X X X X 10 Helmut Schmidt 11 0 X X X X 10 Helmut Schmidt 11 0 X X X X 10 Helmut Schmidt 11 0 X X X X 11 0 X X X X X X 12 Helmut Schmidt 11 0 X X X 13 14 X X X X X X 14 Helmut Schmidt 11 0 X X X 15 16 Helmut Schmidt 11 0 X X 15 16 Helmut Schmidt 11 0 X X 15 16 Helmut Schmidt 11 0 X X 16 Helmut Schmidt Hel	6 US presidential election 2016	13	\checkmark	-	-	\checkmark	\checkmark	0	×		mc2 (which has been dubbed "the v		
8 No Olympics 15 Image: Comparison of the isophotoelectric effect. 9 Myanmar general elections 15 Image: Comparison of the isophotoelectric effect. 10 Helmut Schmidt 11 Image: Comparison of the isophotoelectric effect. 0 Swirtler: Abert Einstein Image: Comparison of the isophotoelectric effect. Abert Einstein Nachen Abert Einstein Image: Comparison of the isophotoelectric effect. 10 Hungers Image: Comparison of the isophotoelectric effect. 11 Image: Comparison of the isophotoelectric effect. Image: Comparison of the isophotoelectric effect. 11 Image: Comparison of the isophotoelectric effect. Image: Comparison of the isophotoelectric effect. 11 Image: Comparison of the isophotoelectric effect. Image: Comparison of the isophotoelectric effect. 11 Image: Comparison of the isophotoelectric effect. Image: Comparison of the isophotoelectric effect. 12 Nather Einstein is basing on of the isophotoelectric effect. Image: Comparison of the isophotoelectric effect. 12 Nather Einstein iso ison of the photoelectric effect. Image: Comparison of the isophotoelectric effect. 13 Nather Einstein iso ison of the photoelectric effect. Image: Comparison of the ison of the photoelectric effect. 14 Neederive and the Nobel Prize on Physics. Image: Comparison of the ison of the photoelectric effect. 14 Neederive and the Nobel Prize on Physics. Image: Comparison of the ison of the photoelectric effect. 14 Neederive and the Nobel Prize on P	7 Refugee Crisis	20	-	-	0	×	×	×	×		1921 Nobel Prize in Physics for his "services to theoretical physics" in		
9 Myanmar general elections 1 10 Helmut Schmidt 1 11 X X 11 X	8 Rio Olympics	15	-	-	0	×	×	×	×	/	particular his discovery of the law of photoelectric effect a pivotal step		
10 Helmut Schmidt 1 Image: Comparison of the second the second the second the second the	9 Myanmar general elections	15	0	×	×	×	×	×	×		the evolution of quantum theory.		
Comparison of the laws of the electromy of the laws of the photosectric effect entry of the laws of the entry of the laws of the photosectric effect entry of the laws of the entry of the laws of the photosectric effect entry of the laws of the entry of the laws of the photosectric effect entry of the laws of the entry of the laws of the entry of entry of the laws of the entry of entry of the laws of the entry of entry of the laws of the photosectric effect entry of entry of the laws of the entry of entry of t	10 Helmut Schmidt	11	0	×	×	×	×	×	×		Near the beginning of his career, Firstein thought that Newtonian		
Albert Einstein was born on 14 March 1879. He was a theoretical physicist. Albert Einstein developed the general theory of relativity. One of his important scientific works deals with the ground-breaking special theory of relativity. He developed the general theory of relativity. Albert Einstein was born as the first child of the Jewish couple Hermann and Pauline Einstein, nee Koch, in Ulm on March 14, 1879 Albert Einstein was born an the Mobel Prize for physics for his explanation of the photoelectric effect. He received the 1921 Nobel Prize in Physics. Being too remote from Sweden, Albert Einstein was quite a good musician. Later, he immigrated to the U.S. [] Allor 6300 Words Ea. 106/300 Words DSWriter: Albert Einstein March 14, Einstein was DSWriter: Albert Einstein March 14, Einstein Was Allor Clapsice and Trevor Lipscombe note that some papela have reported that Einstein was quite a good musician. Later, he immigrated to the U.S. [] DSWriter: Albert Einstein March 14, Einstein Was DSWriter: Albert Einstein March 14, Einstein Was DSWriter: Albert Einstein March 14, Einstein Was DSWriter: Albert Einstein March 14, Einstein Was Allor Clapsice Albert Einstein Was Albert Einstein Was awarded the Stobel Prize Was DSWriter: Albert Einstein March 14, Einstein March 14, Einstein Was Albert Einstein Was Albert Einstein Was DSWriter: Albert Einstein March 14, Einstein March 14, Einstein Was DSWriter: Albert Einstein March 14,	Albert Einstein	osition	А	ll Nuggets			St	ep Comp	olete		3. Best nugget s Group 2		
 Albert Einstein was born as the first child of the Jewish couple Hermann and Pauline Einstein, nee Koch, in Ulm on March 14, 1879 Albert Einstein was born on 14 March 14, 1879 Albert Einstein could not attend the Nobel Prize for physics for his explanation of the photoelectric effect. Albert Einstein was quite a good musician. Later, he immigrated to the U.S. [] Albert Einstein was quite a good musician Ca. 106/300 Words Albert Einstein DSWriter: Albert Einstein Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician. Later, he immigrated to the U.S. [] Albert Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician Cate Calaprice and Trevor Lipscombe and the treve and the Nobel Prize was announced on 9 November 1922. Being too remote from Sweden, Albert Einstein was quite a good musician Cate Calaprice and Calaprice and Cala	Albert Einstein Albert Einstein was born on 14 March 1879. He was a theoretical physicist. Albert Einstein developed the general theory of relativity. One of his important scientific works deals with the ground-breaking special theory of relativity. He is best known for his mass-energy equivalence formula. For his discovery of the law of the photoelectric effect,		A	Il Nugget • Einst equiv • He d	ts tein is be valence f	st knowr formula	for his		He developed the general theory of Group 3				
he received the 1921 Nobel Prize in Physics. Being too remote from Sweden, Albert Einstein could not attend the Nobel Prize Award Ceremony in Stockholm. Alice Calaprice and Trevor Lipscombe note that some people have reported that Einstein was quite a good musician. Later, he immigrated to the U.S. [] Ca. 106/300 Words DSWriter: Albert Einstein DSWriter: Albert Einstein DSWriter: Albert Einstein Pashboard User: Demo (Logout) MDSWriter: Albert Einstein				 Albert Einstein was born as the first child of the Jewish couple Hermann and Pauline Einstein, nee Koch, in Ulm on March 14, 1879 Albert Einstein was born on 14 March 1879 In 1921, he won the Nobel Prize for physics for his 							Albert Einstein was born as the first March 14, 1879 Albert Einstein was born on 14 Marc		
people have reported that Einstein was quite a good musician. Later, he immigrated to the U.S. [] The Nobel Prize was announced on 9 November 1922 The Nobel Prize was announced on 9 November 1922 Being too remote from Sweden, Albert Einstein could not attend the Nobel Prize Award Ceremony in Stockholm One of them deals with the ground-breaking special theory of relativity Some people have reported that Einstein was quite a good musician Albert Einstein was announced on (Step 4: Co MDSWriter: Albert Einstein MDSWriter: Albert Einstein	he received the 1921 Nobel Prize in Phys remote from Sweden, Albert Einstein cou the Nobel Prize Award Ceremony in Stoc Alice Calaprice and Trevor Lipscombe no	tics. Being too uld not attend kholm. te that some	 He received the 1921 Nobel Prize in Physics forhis discovery of the law of the photoelectric effect Albert Einstein was awarded the 1921 Nobel Prize 								In 1921, he won the Nobel Prize for He received the 1921 Nobel Prize in		
 One of them deals with the ground-breaking special theory of relativity. Some people have reported that Einstein was quite a good musician DSWriter: Albert Einstein 	musician. Later, he immigrated to the U.S. []		 The Nobel Prize was announced on 9 November 1922 Being too remote from Sweden, Albert Einstein could not attend the Nobel Prize Award Ceremony in Stockholm 								<i>from</i> Wikipedia: Albert Einstein: (which has been dubbed "the world" "services to theoretical physics", in the evolution of quantum theory. N		
ca. 106/300 Words • Some people have reported that Einstein was quite a good musician • The Nobel Prize was announced o (Step 4: Co MDSWriter: Albert Einstein • Dashboard User: Demo (Logout) MDSWriter: Albert Einstein				One spect	of them ial theor	deals wit y of relat	h the gro ivity	ound-bre	aking		Albert Einstein was awarded the 19		
ADSWriter: Albert Einstein Dashboard User: Demo (Logout) MDSWriter: Albert Einstein	ca. 106/300 Words			• Some	e people	1		at Einste	in was				
ADSWriter: Albert Einstein Dashboard User: Demo (Logout) MDSWriter: Albert Einstein				quite	e a good	nave rep musician	orted th		-		The Nobel Prize was announced on		
				quitte	e a good	nave rep musician	orted th		•		The Nobel Prize was announced on (Step 4: Co-		
	ADSWriter: Albert Einstein 6. Summary organi	zation			e a good	nave rep musician	orted th ashboard St	User: ep Comp	Demo (Logout) plete		The Nobel Prize was announced on (Step 4: Co- MDSWriter: Albert Einstein 5. Sentence form		

ADSWriter: Albert Einstein 1. Nugget identificat	ion	Dashboard User: Demo (Logout) Step Complete
Wikipedia: Albert Einstein	Quotes about Einstein No document selected Wikipedia: Albert Einstein	▼ [?] Einstein is best known [3] X
Wikipedia: Albert Einstein Albert Einstein was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside	Nobelprice.org: Questions and Answers General relativity Einstein family Biography.com: Albert Einstein Einstein for kids: Biography Quotes about Einstein	<pre> for his mass-energy equivalence formula [3]</pre>
quantum mechanics). Einstein's work is also known for its influence on the philosophy of science. Einstein is best	Mass-energy equivalence Annus Mirabilis papers History of physics	[?] He received the 1921 Nobel Prize in Physics for [2] & X
known in popular culture for his mass-energy equivalence formula E = mc2 (which has been dubbed "the world's most famous equation"). He received the	Some people have reported that Ei was quite a good musician, but oth weren't so enthusiastic. A profession violinist claimed be "fiddled like a	nstein ers nal photoelectric

Summarv

MDS Writer is an open-licensed software for designing complex annotation setups with multiple steps. It is particularly useful for the creation of summarization corpora.

Highlights:

- Divide complex annotation setups into multiple steps
- Support cross-document tasks such as multi-document summarization
- Link annotation tool and guidelines to support human annotators in creating high-quality corpora
- Separately store results of intermediate steps and system–user interaction
- Improve evaluation of automatic summarization methods by assessing intermediate results
- First step towards our vision for nextgeneration summarization systems that learn the human summarization process rather than only replicating its *result*



s for his explanation of the photoelectric effect

discovery of the law of the photoelectric effect

st famous equation"). He received the 1921 Nobel Prize in Physics for his cular his discovery of the law of the photoelectric effect, a pivotal step in

obel Prize in Physics in 1922

ember 1922

ference resolution)

Modified nugget

Ilation

inginal nagget (which has been dubbed "the world's most fai effect [2] 🖉 🗙 [Alice Calaprice and Trevor Lipscombe] Some people have eported that instein was quite a good musician 🔗 🗙

Easy to extend to other summarization setups and other cross-document tasks

- Available as open-source software from GitHub under the Apache License (ASL)
 - https://github.com/UKPLab/mdswriter

Extensibility

- MDS *Writer* is flexible to deviate from our prototypical annotation workflow:
 - Define the steps of your own workflow, create and test annotation guidelines
 - Add, modify, or reorder existing steps
 - Extend i18n if necessary (currently: en/de)
- Enables a wide range of complex annotation setups such as:
 - Summarization (single- and multi-document, structured and opinionated summaries,...)
- Information extraction (separate steps for entity, event, relation identification)
- Text compression (sentence fusion, subclause)

≥w	\equiv Albert Einstein was born on 14 March 1879.	▼	
oup se :drop)	\equiv Albert Einstein was a theoretical physicist.	▼	
	∃ Achievements	e×.	
	\equiv Albert Einstein developed the general theory of relativity.	▼	E
	\equiv One of the scientific works deals with the ground-breaking special theory of relativity.	▼	
	\equiv Einstein is best known for his mass-energy equivalence formula.	▼	
	∃ Nobel Prize	e×.	
	\equiv Albert Einstein received the 1921 Nobel Prize in Physics for his discovery of the law of the photoelectric effect.	•	
	from Wikipedia: Albert Einstein:		
	(which has been dubbed "the world's most famous equation"). Albert Einstein received the 1921 Nobel Prize in Physics for his discovery of the law of the photoelectric effect. his "services to theoretical physics", in particular undefined, a pivotal step in the evolution of quantum theory. No	ear	

(which has been dubbed the world's most ramous
equation").
He received the 1921 Nobel Prize in Physics for his
"services to theoretical physics", in particular
his discovery of the law of the photoelectric effect, a

pivotal step in the evolution of quantum theory. Near

Albert Einstein received the 1921 Nobel Prize in Physics

for his discovery of the law of the photoelectric effect.

Previous nugget Nugget 6 of 9 Next nugget

physicist. He developed the general theory of relativity, one of the two pillars of modern physics (alongside quantum mechanics). Einstein's work is also known for its influence on the philosophy of science. Einstein is best known in popular culture for his mass-energy equivalence formula E = mc2 (which has been dubbed "the world's most famous equation"). He received the 1921 Nobel Prize in Physics for his "services to theoretical physics", in particular his discovery of the law of the photoelectric effect, a pivotal step in the evolution of quantum theory.

Albert Einstein was a German-born theoretical

Wikipedia: Albert Einstein

Dashboard | User: Demo (Logout)

Step Complete

Near the beginning of his career, Einstein thought that Newtonian mechanics was no longer enough to reconcile the laws of classical mechanics with the laws of the electromagnetic field. This led to the development of his special theory of relativity. He realized, however, that the principle of relativity could also be extended to gravitational fields, and with his subsequent theory of gravitation in 1916, he published a paper on general relativity. He continued to deal with problems of statistical mechanics and quantum theory, which led to his explanations of particle theory and the motion of molecules. He also investigated the thermal properties of light which laid

removal, co-references, lexical substitution,...)

Cross-document discourse structure annotation

Technology

- Java/JSP and JavaScript
- Multi-user, multi-step, multi-doc. support
- Individual user actions sent to server application in real-time based on WebSockets connection
- Results stored in SQL database
- Simple text-based protocol allows for easy and fast extensions

This work has been supported by the DFG-funded research training group "Adaptive Preparation of Information from Heterogeneous Sources" (AIPHES, GRK 1994/1) and by the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant No. I/82806.

https://www.aiphes.tu-darmstadt.de

MDS *Writer* – https://github.com/UKPLab/mdswriter Source code, documentation, annotation guidelines, video tutorial