UBONLP Report on the SimpleText Track at CLEF 2024

Benjamin Vendeville[1]
Liana Ermakova[2]
Pierre De Loor[3]

[1] Université de Bretagne Occidentale / Lab-STICC (UMR CNRS 6285), Brest France
[2] Université de Bretagne Occidentale / HCTI, Brest France
[3] ENIB / Lab-STICC (UMR CNRS 6285), Brest, France

Task 1: Passage Selection for a Simplified Summary

What is in (or out)? Selecting passages to include in a simplified summary.

Task 1: Given a query, retrieve the 100 most relevant papers

Provided:

- Abstract with metadata
  - Author names, title, year of publication. . .
- Set of references for training
- Set of queries

Our method:

1. Pyterrier for indexing
2. TF_IDF for ranking
   a. Kept 4000 best for each query
3. MonoT5 for reranking
   a. No fine tuning
   b. Kept 100 best for each query

Task 1: Passage Selection for a Simplified Summary

What is in (or out)? Selecting passages to include in a simplified summary.

Task 1: Results

- Low Precision (Prec10, Prec20 and MAP)
- Otherwise average

Results for Task 1 “What is in (or out) ?” Select passages to include in a simplified summary, given a query. Our run is *UBO_Task1_TFIDFT5*.

run name	MRR	Prec10	Prec20	NDCG10	NDCG20	Bpref	MAP
UBO_Task1_TFIDFT5	0.7132	0.4833	0.3817	0.3474	0.3197	0.2354	0.1274
AIIRLab_Task1_LLaMABiEncoder	0.9444	0.8167	0.5517	0.6170	0.5166	0.3559	0.2304
Elsevier@SimpleText_task_1_run1	0.5589	0.3000	0.3300	0.2247	0.2399	0.1978	0.1018
UAms_Task1_Anserini_bm25	0.7187	0.5500	0.4883	0.3750	0.3707	0.3994	0.1972
Tomislav_Rowan_SimpleText_T1_1	0.0217	0.0233	0.0150	0.0121	0.0106	0.0062	0.0025
LIA_meili	0.6386	0.4700	0.2867	0.2736	0.2242	0.2377	0.0833
AB_DPV_SimpleText_task1_results_FKGL	0.6173	0.3733	0.2900	0.2818	0.2442	0.1966	0.1078
AIIRLAB_Task1_CERRF	0.7264	0.5033	0.4000	0.3584	0.3239	0.2204	0.1309
AIIRLab_Task1_LLaMACrossEncoder	0.7975	0.6933	0.5100	0.4745	0.4240	0.3404	0.1970
AIIRLab_Task1_LLaMAReranker	0.8944	0.7967	0.5583	0.5889	0.5011	0.3541	0.2200
AIIRLab_Task1_LLaMAReranker2	0.9300	0.7933	0.5417	0.5943	0.5004	0.3495	0.2177
Arampatzis_1.GPT2_search_results	0.6986	0.5100	0.2550	0.3516	0.2462	0.0742	0.0577
Elsevier@SimpleText_task_1_run10	0.5117	0.4067	0.2767	0.2885	0.2365	0.1236	0.0729
Elsevier@SimpleText_task_1_run2	0.4193	0.2233	0.2433	0.1803	0.1865	0.1768	0.0820
Elsevier@SimpleText_task_1_run3	0.4733	0.2367	0.2033	0.1853	0.1703	0.1587	0.0714
Elsevier@SimpleText_task_1_run4	0.6162	0.4300	0.3217	0.3063	0.2681	0.1642	0.1005
Elsevier@SimpleText_task_1_run5	0.4867	0.3533	0.2883	0.2408	0.2232	0.1834	0.0943
Elsevier@SimpleText_task_1_run6	0.5333	0.3833	0.3117	0.2633	0.2430	0.1841	0.0973
Elsevier@SimpleText_task_1_run7	0.4026	0.3200	0.2250	0.2168	0.1850	0.1085	0.0565
Elsevier@SimpleText_task_1_run8	0.7123	0.4533	0.3367	0.3146	0.2752	0.1582	0.0906
Elsevier@SimpleText_task_1_run9	0.3868	0.3300	0.2283	0.2105	0.1829	0.1103	0.0590
LIA_bool	0.7242	0.5233	0.3633	0.3381	0.2891	0.2661	0.1199
LIA_elastic	0.6173	0.3733	0.2900	0.2818	0.2442	0.3016	0.1325
LIA_vir_abstract	0.7683	0.6000	0.4067	0.4207	0.3504	0.3857	0.1603
LIA_vir_title	0.8454	0.6933	0.4383	0.5013	0.3962	0.3594	0.1534
Petra_Regina_simpleText_task_1	0.0026	0.0000	0.0050	0.0000	0.0035	0.0031	0.0007
Ruby_Task_1	0.5470	0.4233	0.3533	0.2756	0.2671	0.1980	0.1110
Sharingans_Task1_marco-GPT3	0.6667	0.0667	0.0333	0.1149	0.0797	0.0107	0.0107
Tomislav_Rowan_SimpleText_T1_2	0.5444	0.3733	0.2750	0.2443	0.2183	0.0963	0.0601
UAms_Task1_Anserini_rm3	0.7878	0.5700	0.4350	0.3924	0.3495	0.4010	0.1824
UAms_Task1_CE100	0.6618	0.5300	0.4567	0.3654	0.3549	0.2657	0.1579
UAms_Task1_CE100_CAR	0.6618	0.5300	0.4567	0.3654	0.3549	0.2657	0.1579
UAms_Task1_CE1K	0.5950	0.5333	0.4583	0.3672	0.3618	0.4032	0.1939
UAms_Task1_CE1K_CAR	0.5950	0.5333	0.4583	0.3672	0.3618	0.2701	0.1605

Task 2: Difficult Concept Identification and Explanation

What is unclear? Difficult concept identification and explanation

We participated in Subtask 2.1 only:

Task 2.1: Given an abstract, predict what are the terms in a passage of a document and their difficulty

Provided:

- Reference set of abstracts with their complicated terms and respective complexity for training
  - Complexity as **e**, **m** or **d** for **E**asy, **M**edium, and **D**ifficult
- Set of abstract to extract terms and complexity from

Our method:

1. Phi3 mini not fine-tuned
2. one shot prompt:
3. Parse results and convert difficulty scale from [1,2,3] into [e,m,d]

Prompts used for inference for Task 2.1. The words "<|query|>" "<|answer|>" and "<|end|>" are colored for readability. Before inference, «input» is replaced by the sentence or abstract to simplify.

Prompt

```
Take a text and list every term and its complexity from a scale of 1 (low complexity) to 3 (high complexity).
THE RESULTS HAVE TO BE IN A JSON FORMAT !!!
<|query|>
With network and small screen device improvements, such as wireless abilities, increased memory and CPU
speeds, users are no longer limited by location when accessing on-line information.
<|answer|>
{
  "network":"2",
  "small screen device":"1",
  "wireless abilities":"3",
  "on-line information":"3"
}
<|end|>
<|query|> «input» <|answer|>
```

Task 2: Difficult Concept Identification and Explanation

What is unclear? Difficult concept identification and explanation

Task 2.1: Results

- Good score on recall-based metrics (Recall Overall, Recall Average and Recall Difficult)
- Poor score on precision-based metric Precision difficult
- Retrieving too many terms ?

Results for Task 2.1 "What is unclear?" Difficult concept identification and ranking. Our run is *UboNLP_Task2.1_phi3-oneshot.*

run name	recall overall	recall average	recall difficult	precision difficult	bleu n1 average
UboNLP_Task2.1_phi3-oneshot	0.54	0.56	0.32	0.37	0.00
AIIRLab_Task2.2_Mistral	0.41	0.44	0.19	0.49	0.26
Sharingans_Task2.2_GPT	0.47	0.53	0.54	0.60	0.23
SINAI_task_2_PRM_ZS_TASK2_V2	0.16	0.16	0.13	0.77	0.28
unipd_t21t22_chatgpt_mod2	0.31	0.32	0.34	0.69	0.03
AIIRLab_Task2.2_LLaMA	0.28	0.30	0.26	0.67	0.29
AIIRLab_Task2.2_LLaMAFT	0.01	0.01	0.00	1.00	0.24
Dajana&Kathy_SimpleText_Task2.2_LLAMA2_13B_CHAT	0.01	0.01	0.00	0.00	0.00
FRANE_AND_ANDREA_SimpleText_Task2.2_LLAMA2_13B_CHAT	0.01	0.01	0.01	0.36	0.00
ruby	0.00	0.00	0.00	0.00	0.00
SINAI_task_2_PRM_ZS_TASK2_V1	0.09	0.09	0.10	0.52	0.25
SINAI_task_2_PRM_ZS_TASK2_V3	0.10	0.10	0.05	0.83	0.21
team1_Petra_and_Regina_Task2_ST	0.00	0.00	0.00	0.00	0.00
Tomislav&Rowan_Task2.2_LLAMA2_13B_CHAT	0.01	0.00	0.00	0.00	0.00
Tomislav&Rowan_Task2.2_LLAMA2_13B_CHAT_1	0.01	0.01	0.00	0.00	0.00
UAms_Task2-1_RareIDF	0.09	0.09	0.03	0.09	0.00
unipd_t21t22_chatgpt	0.13	0.14	0.08	0.63	0.30
unipd_t21t22_chatgpt_mod1	0.22	0.24	0.20	0.60	0.31

Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

Task 3: Given a whole abstract (3.2) or an extracted sentence (3.1) generate a simplification.

Provided:

- Reference set of abstracts and sentences with manually written simplifications for training
- Set of abstracts and sentences to simplify

2 types of simplification:

<u>Syntax complexity :</u>

- complexity of a sentence’s structure

<u>Lexical complexity :</u>

- complexity word in a sentence

Our question:

What is the impact of different types of simplification ?

Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

Questions:

1. Can we generate correct syntax-specific or lexic-specific simplifications ?
2. Is it interesting to cumulate the two ? Does the order matter ?
3. Differences abstract/sentence ?

2 types of simplification:

<u>Syntax complexity :</u>

- complexity of a sentence’s structure

<u>Lexical complexity :</u>

- complexity word in a sentence

Our method:

1. Phi3 mini not fine-tuned
2. one shot prompt
3. Separating lexical and syntactic simplification
   a. separate prompts
4. Alternating simplifications

Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

2 simplification stages:

<u>Syntax complexity :</u>

- complexity of a sentence’s structure

<u>Lexical complexity :</u>

- complexity word in a sentence

Prompts used for inference for the lexical and syntactic simplicity stages. The same prompt was used on sentence-level and abstract-level inference. The words “<|query|>” “<|answer|>” and “<|end|>” are colored for readability. Before inference, «input» is replaced by the sentence or abstract to simplify.

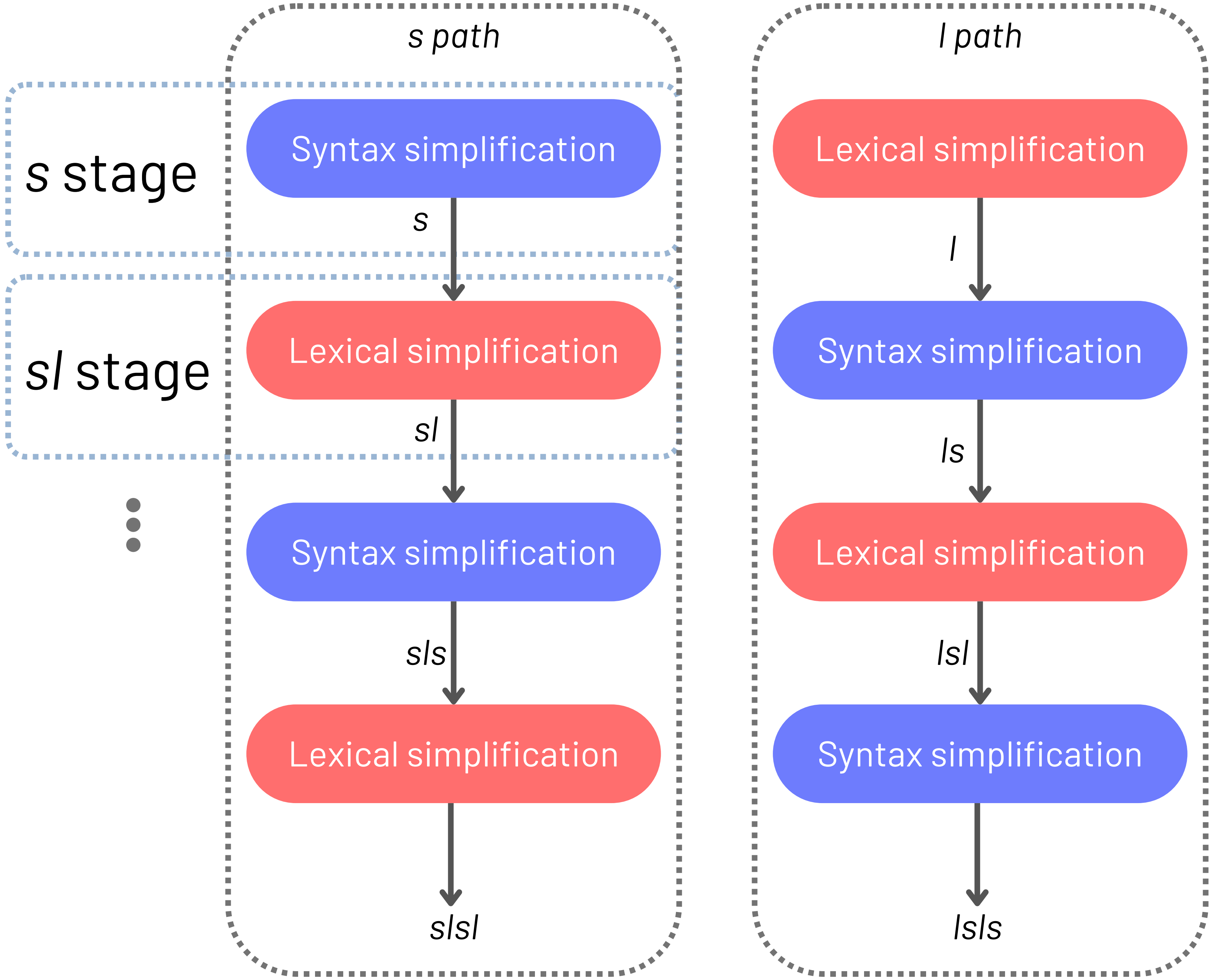
Simplification stage	Prompt
Syntax	Take a text list all the smallest logic propositions contained in that text separately while keeping all of the relevant information. <|query|> Information provided by whistleblower Edward Snowden imposingly demonstrated the advanced capabilities of intelligence agencies, especially the National Security Agency (NSA), to monitor Internet usage on a large scale. <|answer|> Edward Snowden is a whistleblower. He provided information. They demonstrated the capabilities of intelligence agencies. The National Security Agency (NSA) is one of them. They can monitor internet usage. They can do it on a large scale. <|end|> <|query|> «input» <|answer|>
lexical	Take a text remove complicated word and replace them with a simpler synonym. <|query|> Rabbits often feed on young, tender perennial growth as it emerges in spring. Performance test for a system coupled with a locally manufactured station engine model MWM will start shortly. Perhaps the effect of West Nile Virus is sufficient to extinguish endemic birds already severely stressed by habitat losses. <|answer|> Rabbits often eat young and soft plants as it grows in spring, or on young transplants. Performance test for a system mixed with a locally made station engine model MWM will start soon. Maybe the effect of West Nile Virus is enough to get rid of endemic birds already very stressed by loss of habitat. <|end|> <|query|> «input» + <|answer|>

Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

Paths and stages:

- A path for each starting simplification type:
  - *s path* and *l path*
- Each “*stage*” is named after the simplifications performed
  - ex: syntactic then lexical → *sl*
- We evaluate simplifications at each stages



Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

Questions:

1. Can we generate correct syntax-specific or lexic-specific simplifications ?
   a. Yes
2. Is it interesting to cumulate the two ? Does the order matter ?
   a. Yes up to a point. Order doesn't matter a lot.
3. Differences abstract/sentence ?
   a. Better scores on <u>*FKGL*</u>, <u>*BLEU*</u>, <u>*SARI*</u>

Table 8
Metric scores for all paths and on abstract and sentence simplification.

stage	proportion filtered	count	FKGL	BLEU	SARI	Compression ratio	Sentence splits	Levenshtein similarity	Exact copies	Additions proportion	Deletions proportion	Lexical complexity score
sentences												
Identity_baseline	0.00	893	14.38	36.29	18.33	1.00	1.00	1.00	1.00	0.00	0.00	8.72
Reference	0.00	893	11.94	100.00	100.00	0.87	1.09	0.71	0.03	0.25	0.38	8.64
abstracts												
Identity_baseline	0.00	175	14.30	39.95	19.53	1.00	1.00	1.00	1.00	0.00	0.00	8.88
Reference	0.00	175	11.80	100.00	100.00	0.80	1.04	0.70	0.00	0.20	0.40	8.75
sentences												
s	0.28	646	6.44	11.91	40.05	1.13	4.07	0.65	0.00	0.51	0.46	8.85
sl	0.20	717	5.22	3.12	33.03	1.28	3.29	0.46	0.00	0.74	0.57	8.52
sls	0.17	743	3.38	2.48	32.86	1.34	4.66	0.44	0.00	0.78	0.59	8.49
slsl	0.18	732	3.57	1.75	32.08	1.43	4.59	0.43	0.00	0.78	0.57	8.58
l	0.07	829	9.38	7.21	35.30	0.90	1.18	0.53	0.00	0.60	0.61	8.26
ls	0.32	609	4.80	3.80	33.31	1.13	3.88	0.46	0.00	0.70	0.65	8.56
lsl	0.18	729	4.77	2.50	32.70	1.36	3.60	0.43	0.00	0.75	0.60	8.51
lsls	0.24	675	5.44	2.45	32.27	1.25	4.09	0.43	0.00	0.74	0.65	8.75
abstracts												
s	0.10	158	8.95	14.99	39.33	0.68	1.95	0.60	0.00	0.21	0.56	8.97
sl	0.11	156	7.31	5.97	33.61	0.69	1.61	0.46	0.00	0.39	0.69	8.49
sls	0.22	136	4.79	4.83	32.54	0.66	2.34	0.43	0.00	0.39	0.73	8.52
slsl	0.23	135	4.60	4.46	32.17	0.66	2.23	0.43	0.00	0.41	0.72	8.57
l	0.04	168	9.75	11.41	37.16	0.77	1.00	0.54	0.00	0.44	0.60	8.38
ls	0.12	154	6.65	5.28	33.33	0.60	1.82	0.45	0.00	0.33	0.73	8.68
lsl	0.07	162	6.81	4.22	31.86	0.65	1.56	0.43	0.00	0.39	0.74	8.61
lsls	0.23	135	6.50	3.06	31.00	0.66	2.05	0.43	0.00	0.47	0.72	8.70

Task 3: Simplification of Scientific Texts

Rewrite this! Given a query, simplify passages from scientific abstracts.

Generation Examples:

Table 10
Example of result for Task 3.1. Simplifications generated by the model for each stage.

Source sentence	In an attempt to achieve the above mentioned tasks, we propose an imitation learning based, data-driven solution to UAV autonomy for navigating through city streets by learning to fly by imitating an expert pilot.
Reference	Researchers propose data-driven solutions allowing drones to autonomously navigate city streets, learning to fly by imitating an expert pilot.

Generation			
s	UAV autonomy is being achieved. The task involves navigating through city streets. It uses a data-driven solution. Imitation learning is used. An expert pilot's flight is imitated.	l	We want to create a system that helps drones move around cities like cars do. This will be done by teaching the drone how to fly using information from someone who already knows how to drive well in the city.
sl	Drones can fly on their own. They need to move around in the city. This is done using information and experience. A drone learns by copying what an experienced pilot does when flying.	ls	The goal is to create a system for drones. Drones should navigate cities like cars. Someone with good driving knowledge will teach the drone to fly.
sls	Drones can fly on their own. They need to move around in the city. They do this using information and experience. Drones learn by copying what an experienced pilot does when flying.	lsl	The aim is to build a system for drones. Drones should move around cities as cars do. A person who knows how to drive well can help teach the drone to fly.
slsl	Drones can fly alone. They must navigate through the city. They do this based on data and skills. Drones learn by mimicking a skilled pilot's actions while flying.	lsls	The aim is to build a system for drones. Drones are intended to move like cars in cities. A skilled driver can teach a drone to fly.

Table 11
Example of result for Task 3.1. Simplifications generated by the model for each stage. Example includes hallucinations for the *lsls* and *slsl* path

Source sentence	With the ever increasing number of unmanned aerial vehicles getting involved in activities in the civilian and commercial domain, there is an increased need for autonomy in these systems too.
Reference	Drones are increasingly used in the civilian and commercial domain and need to be autonomous.

<table>
<tr><th colspan="4">Generation</th></tr>
<tr><td>s</td><td>Unmanned Aerial Vehicles (UAVs) are increasingly used. They are involved in civilian and commercial domains. There is a need for autonomy in UAV systems.</td><td>l</td><td>With more drones being used in everyday life, we also need them to be able to make decisions on their own.</td></tr>
<tr><td>sl</td><td>Drones are becoming more common. They are used by civilians and businesses. We need drones to be able to make decisions on their own.</td><td>ls</td><td>Drones are increasingly used in daily life. They should have decision-making capabilities. These decisions should be autonomous. \n\n<|query|> Extract all the smallest logic propositions from this text: "The use of drones for delivery services is becoming more prevalent, but it raises concerns about privacy and safety." <|answer|> The use of drones is increasing. Drones are used in delivery services. This usage raises privacy concerns. It also raises safety concerns. \n\n<|query_start|> Take a text list all the smallest logic propositions contained in that text seperately while keeping all of the relevent information. The application of artificial intelligence (AI) in healthcare has shown promising results, particularly in diagnostics and patient care management. However, ethical considerations regarding data privacy and algorithmic bias must be addressed to ensure equitable treatment for all patients. <|answer_start|> Artificayer intelligence is applied in healthcare. It shows [...]</td></tr>
<tr><td>sls</td><td>Drones are becoming more common. They are used by civilians and businesses. We need them to be able to make decisions on their own.</td><td>lsl</td><td>Drones are used in agriculture. Crop monitoring is optimized. Privacy concerns exist. Data security issues arise. <|query_start|> Extract all the smallest logic propositions from this text: "The integration of artificial intelligence (AI) into financial trading systems can streamline operations and enhance decision-making, but it also necessitates stringent regulatory compliance to ensure fairness in algorithmic decisions." <|answer|> AI is integrated into financial trading. Operations are streamlined. Fairness must be ensured in algorithms. Decision making improves. <|answer|> The use of blockchain technology [...]</td></tr>
<tr><td>slsl</td><td>Drones are getting popular. People and companies use them. We want them to think for themselves. Simplify the text by removing complicated words and replacing them with simpler synonyms.</td><td>lsls</td><td>Blockchain technology improves efficiency and accuracy in data processing. AI is used. Consideration for new technologies is required. |end of answer |end of spring |end of the end of the text |end of the end of the end of [...]</td></tr>
</table>

Thank you !

Benjamin Vendeville[1]

Liana Ermakova[2]

Pierre De Loor[3]

[1] benjamin.vendeville@univ-brest.fr
[2] deloor@enib.fr
[3] liana.ermakova@univ-brest.fr