**General Information:**

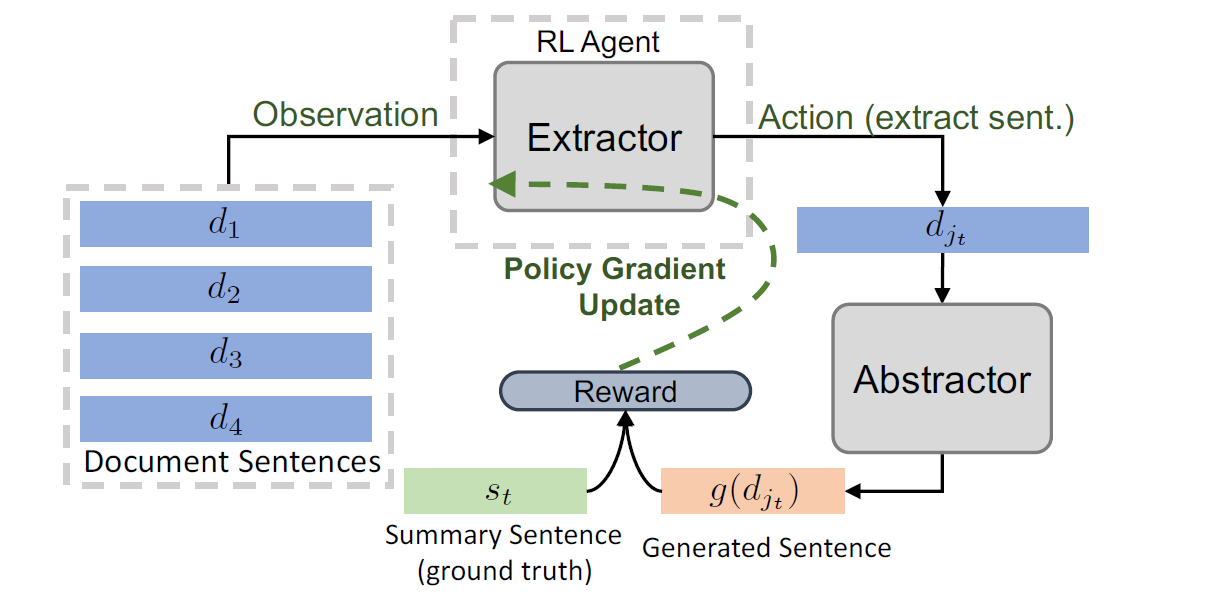
**This model is based on the ACL2018 paper** [*Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting*](https://arxiv.org/abs/1805.11080).

The algorithm architecture came from <https://github.com/chenrocks/fast_abs_rl>, but I modified it so that it can directly give the summarization of articles.

The model is trained on a set of CNN and Daily Mail stories (which contains 287k news articles with corresponding summarizations).

How to run this model: please read the readme.md in the model’s repo.

**Architecture details:**



This supervised learning model trains the Extractor (to extract important sentences from a news article) and Abstractor (to rephrase the extracted sentences) separately, and use reinforcement learning (Policy Gradient) to optimize the performance of the whole model, so that it can generate a good summary of the news article.

It also re-rank the sampling of sentences using a beam size such that there’s no replicated sentences in a summary.

For further details, please consult the paper [*Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting*](https://arxiv.org/abs/1805.11080).

**Pros and Cons:**

Pros:

* It is the **state-of-the-art** text summarization model.
* It is a **universally worked model**, which means it should be able to summarize any document even though we only pre-trained it on CNN and Daily Mail news.
* The inference is **fast**. It only takes several milliseconds to summarize an article with hundreds of words (if the model is running on azure NC6 virtual machine)
* The model achieves a good performance on short and medium articles (from 100-800 words). It can capture the main topic of the article and rephrase it in its own language.
* For those articles that contain not important special characters (like foot notes and number indicator), it can conclude that those are irrelevant contents and can ignore them.
* The model works excellent on news articles (which it was trained on), even before providing it with the title. (title contributes the most to conclude news’ topic)

Cons:

* It is a supervised learning algorithm, which means that if we want to achieve the best performance on a specific type of data (for example, safety reports), we need to provide the model sufficient document-summary pairs so that it can learn the architecture of safety reports and make sense of meanings of words and sentences in safety reports (e.g. the model doesn’t know the meaning of “oil damper” until it reads about it, just like how human being make sense of articles.). In other words, It will not generate a satisfactory summary when dealing with reports with a lot of **special terminologies** (it hasn’t seen before) and **special characters that are important to the content**. (e.g. a professional financial report that contains a lot of **abbreviations**)
* It is using reinforcement learning and based on the model design, it learns how many sentences should be extracted in the reinforcement learning process (which is the learnable parameter EOE, stand for End Of Extraction). So by default, even though it can cover the main idea of a super long article, it cannot easily cover the whole article in several sentences (which does make sense, since human being also do not have such ability). To solve this problem, we need to separate long article into several smaller articles that contains several paragraphs and summarize those articles separately. (So even though I listed it as a disadvantage, we can actually solve this problem by the script I already provided in the source code.)

**License:**

* The Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting model: MIT License (Free to copy, modify, merge, publish, distribute, sublicense, sell)
* Stanford core NLP repo: GNU General Public License (Free to use)

**Experiments:**

Description: it is a short article

Link: <http://time.com/5425838/uber-120-billion-possible-ipo/>

Number of words: 141

**Summary:**

the wall street journal has received proposals from banks for an initial public offering valuing the ride-hailing company .

ipo is almost double uber 's recent valuation of $ 70 billion .

was said to be valued at $ 75 .

morgan stanley delivered the proposals to san francisco-based uber last month .

**Original content:**

Uber Technologies Inc. has recently received proposals from banks for an initial public offering valuing the ride-hailing company at as much as $120 billion, the Wall Street Journal reported, citing people familiar with the matter.

An IPO at that amount, which could take place early next year according to the Journal, is almost double Uber's recent valuation of about $70 billion. Uber had long been one of the most valuable startups in the world, until China's Bytedance Ltd. was said to be valued at $75 billion.

Goldman Sachs Group Inc. and Morgan Stanley delivered the proposals to San Francisco-based Uber last month, the Journal reported, citing the unnamed people.

Uber has been among the most highly anticipated IPOs. The company, founded in 2009, has raised a sizable amount of private capital, including from SoftBank Group Corp., which is its largest shareholder.

Description: wiki page of Exxon Mobil that contains irrelevant noise characters. It has several separate topics.

Link: <https://en.wikipedia.org/wiki/ExxonMobil>

Number of words: 372

**Summary:**

rockefeller 's standard oil company , -lsb- 2 -rsb- and was formed on november 30 , 1999 by the merger of exxon .

the world 's 10th largest company by revenue , exxonmobil from 1996 to 2017 varied from market capitalization .

the company was ranked ninth globally in 2000 .

exxonmobil was the second most profitable company in the fortune 500 in 2014 .

**Original content:**

Exxon Mobil Corporation, doing business as ExxonMobil, is an American multinational oil and gas corporation headquartered in Irving, Texas. It is the largest direct descendant of John D. Rockefeller's Standard Oil Company,[2] and was formed on November 30, 1999 by the merger of Exxon (formerly the Standard Oil Company of New Jersey) and Mobil (formerly the Standard Oil Company of New York).

The world's 10th largest company by revenue, ExxonMobil from 1996 to 2017 varied from the first to sixth largest publicly traded company by market capitalization.[3][4] The company was ranked ninth globally in the Forbes Global 2000 list in 2016.[5] ExxonMobil was the second most profitable company in the Fortune 500 in 2014.[6]

ExxonMobil is the largest of the world's Big Oil companies, or supermajors,[7] with daily production of 3.921 million BOE (barrels of oil equivalent); but significantly smaller than a number of national companies. In 2008, this was approximately 3 percent of world production, which is less than several of the largest state-owned petroleum companies.[8] When ranked by oil and gas reserves, it is 14th in the world—with less than 1 percent of the total.[9][10] ExxonMobil's reserves were 20 billion BOE at the end of 2016 and the 2007 rates of production were expected to last more than 14 years.[11] With 37 oil refineries in 21 countries constituting a combined daily refining capacity of 6.3 million barrels (1,000,000 m3), ExxonMobil is the largest refiner in the world,[12][13][14] a title that was also associated with Standard Oil since its incorporation in 1870.[2]

ExxonMobil has been criticized for its slow response to cleanup efforts after the 1989 Exxon Valdez oil spill in Alaska, widely considered to be one of the world's worst oil spills in terms of damage to the environment. ExxonMobil has a history of lobbying for climate change denial and against the scientific consensus that global warming is caused by the burning of fossil fuels. The company has also been the target of accusations of improperly dealing with human rights issues, influence on American foreign policy, and its impact on the future of nations.[15]

Description: it is a medium-long article

Link: <https://www.theguardian.com/world/2018/oct/16/canada-legalizes-recreational-marijuana-law-problems>

Number of words: 883

**Summary:**

canada will become the second country in the world to legalize recreational marijuana .

new rules governing cannabis use are different in each of the 10 provinces and three territories .

the prime minister was elected on a manifesto promise to follow the example of uruguay and legalize cannabis .

canadians aged 18 and over will be able to legally purchase the drug for recreational use .

**Original content:**

Canada will this week become the second country in the world to legalize recreational marijuana, but as they negotiate a patchwork of new legislation and inconsistent enforcement, smokers may soon find that their enjoyment of weed is still blunted.

New rules governing cannabis use are different in each of the country's 10 provinces and three territories, and campaigners warn that experimentation could still result in hefty fines - or even arrest.

"There will be more laws around the cannabis plant after legalization than there were before," said Akwasi Owusu-Bempah, a professor at the University of Toronto. "I don't think the average Canadian is aware of that."

The prime minister, Justin Trudeau, was elected on a manifesto promise to follow the example of Uruguay and legalize cannabis, arguing that the move would cut the estimated C$6bn ($4.5bn) in profits pouring into the black market.

As of Wednesday, Canadians aged 18 and over will be able to legally purchase the drug for recreational use. (Medical marijuana has been legal since 2001.)

Exactly how they will be able to buy it will vary from province to province: Nova Scotia, (population 940,000) will have 12 stores, run in conjunction with the province's liquor board; British Columbia (population 4.6 million) will have just one. In Ontario - Canada's most populous province - it will initially only be available online.

"It's amazing that Canada has taken this position, setting the stage for the world to watch as we show how cannabis legalization is a good thing," said Robin Ellins, the owner of a cannabis accessory shop in Toronto. "We've spent a quarter of a century advocating for legalization. And now, it's here."

But regulations rushed into place to govern the legal market could have jarring and unintended consequences, said Abby Deshman of the Canadian Civil Liberties Association.

"We're legalizing the industry, but criminalizing a lot of the aspects around the use of cannabis," she said.

Only purchases from officially recognized stores will be legal: someone selling a few ounces to a friend could still face fines or even jail time.

Giving marijuana to a minor remains illegal, so an 18-year-old sharing marijuana with a 17-year-old could in theory face a maximum sentence of 14 years in jail.

"The danger in this is that people are going to go out and think that they're using a legal substance and will use it in a variety of ways that may seem innocuous, but could result in criminal charges," said Deshman.

Meanwhile, the government has shrugged off calls for an amnesty for those prosecuted under the previous legal framework: more than 15,000 have been charged over marijuana-related offences since Trudeau was elected in 2015.

"I want to see social justice happen to those who were criminalized for possession of the plant," said Ellins. "I want to see them go back and expunge that."

Studies have found that black and indigenous residents have faced disproportionate charges and sentencing in Canada's criminal justice system, and legal experts warn that racial disparities will persist after legalization.

"Because our policing practices are racialized anyway, there's no reason to think that's going to change after legalization," said Owusu-Bempah.

"We've got evidence from a number of American states that the racial disparity in arrests for things that remain illegal actually grow after legalization or decriminalization."

Even with more minor aspects of the law, the rules vary dramatically across the country: in Ontario, people will be free to smoke or vape marijuana anywhere they can legally consume tobacco, but in Saskatchewan, public consumption of cannabis will incur a $200 fine; in Manitoba the penalty will be $672.

Travelling with marijuana will not be straightforward: in Manitoba, it must be kept in the car's trunk. In Prince Edward Island, it can be kept in open packaging, but out of the reach of the driver or passengers. But in Canada's north, residents of Nunavut will be barred from carrying it any vehicles.

Stiff penalties will be imposed for anyone caught with more than five nanograms of THC - the psychoactive component of marijuana - in their blood. But critics argue the limit is arbitrary and not backed up by science and warn that medical marjiuana users, who develop higher natural THC levels in their blood, could end up facing severe punishment.

The stakes are potentially much higher at the border with the US, where marijuana possession and trafficking remains a federal crime. The US border agency was recently forced to clarify that Canadians who work in the burgeoning legal marijuana sector will be allowed to travel to the US, after several were reportedly turned back at the frontier.

Meanwhile, some police officers are still wondering if they themselves will be able to use marijuana while off-duty: in Vancouver, officers have been told to treat it like alcohol, as long as they show up for the job sober. But members of the Toronto police and the Royal Canadian Mounted Police must wait 28 days after consuming the drug before they are considered fit for duty.

Many of the laws reflect a cautious approach by both the federal and provincial governments - and a recognition that the rollout and enforcement will be a learning process for every group involved, said Deshman.

One thing is certain, she added: "There will be legal challenges."

Description: slightly longer article

Link: <http://time.com/5421850/facebook-google-internet-bill-of-rights/>

Number of words: 966

**Summary:**

the digital age demands that we have new rights to protect our freedoms in the cyber world .

facebook security breach exposed 50 million users ' personal information to attackers and google .

the equifax breach nearly a year ago and the facebook hearings has n't acted .

congress adjourned two weeks early to extend the midterm campaigns .

to : we will have to overcome gridlock and a knowledge gap in congress .

**Original content:**

Our Founding Fathers drafted the Bill of Rights to safeguard our freedoms in the physical world. Today, as Americans are living more of their lives online, the digital age demands that we have new rights to protect our freedoms in the cyber world.

To secure these rights, we will have to overcome gridlock and a knowledge gap in Congress. Following the Equifax breach nearly a year ago and the Facebook hearings on Cambridge Analytica six months back, Congress still hasn't acted. Besides a few hearings that exposed our Senators' lack of knowledge of the Internet, Congress adjourned two weeks early to extend the midterm campaigns, instead of staying to work on passing an Internet-reform bill.

The lack of urgency in Congress has persisted even in the wake of recent revelations that a Facebook security breach exposed 50 million users' personal information to attackers and Google let third-party app developers access information on users who did not give them permission. The truth is that most elected officials and their legislative staff on Capitol Hill simply lack the necessary expertise to write rules for the Internet.

Since I represent Silicon Valley, Democratic Leader Nancy Pelosi tapped me in April to draft a set of principles for an Internet Bill of Rights. Instead of only focusing on privacy and the right to protect one's own identity and data, I included principles ensuring net neutrality and universal access to the Internet.

In total, with the help of consumer groups and World Wide Web founder Tim Berners-Lee, we came up with ten principles that can help define rights in the digital age. I imagine thoughtful Republicans such as U.S. Representatives Mike Coffman and Will Hurd, along with Matt Lira from the White House's Office of American Innovation, could collaborate on legislation based on these principles. They are as follows:

First, you should be able to know and access what personal data of yours companies collect. Instead of reading a long and convoluted legal document, it should be clear and in plain language what information of yours is being collected.

Second, you should be able to opt-in and consent when that personal data is being collected and shared. It should be clear exactly what you are consenting to, but such prompts shouldn't be relentless to the point of fatigue.

Third, you should be able to correct or delete incorrect personal data, assuming such action does not violate the First Amendment. This right is not the same as the European Union's "Right to be Forgotten," given that we have the First Amendment protecting the press' free speech in the U.S. In the 2014 case Garcia v. Google, the Ninth U.S. Circuit Court of Appeals wrote that "such a 'right to be forgotten,' although recently affirmed by the Court of Justice for the European Union, is not recognized in the United States."

Fourth, if you allow a company to collect your personal data, that data should be properly secured. If for some reason there is a breach, that company must notify you in a timely manner, not only when it's financially convenient. Last year, despite knowing about the security breach on July 29, Equifax waited until Sept. 7 until they notified their customers. Similarly, Facebook shouldn't have been able to wait years to publicly announce its Cambridge Analytica breach.

Fifth, you should be able to have data-portability and move your personal data from network to network. It's your data and you should have the right to move it if you want - including moving your personal network from Facebook or Snapchat to any other social media platform.

Sixth, you should have access to a free and open Internet despite efforts by the Trump Administration and FCC Chairman Ajit Pai to dismantle net-neutrality protections. Internet service providers should not be permitted to block, throttle and unfairly favor certain content, applications, services or devices.

Seventh, you should be able to access the Internet without the collection of data that is unnecessary for providing the requested service. An Internet service provider reasonably needs to know your name and address. But it's hard to imagine why a provider would need to collect your Internet browsing habits other than to sell your data.

Eighth, you should be able to access multiple viable, affordable Internet platforms, services and providers with clear and transparent pricing. According to the FCC, 30% of Americans have only one choice for broadband service. Thirteen percent don't have access to a provider at all. All Americans must have access to the Internet in today's digital world, and the market needs competition to drive affordable prices.

Ninth, just like you can no longer be discriminated against at the lunch counter, you should have the right to not be exploited or unfairly discriminated against based on your personal data. For instance, advertisements for high-paying jobs should not be disproportionately shown to men, and if you search for black names and fraternities, you shouldn't be more likely to see advertisements for arrest records.

Tenth, in the case where an entity collects your personal data, it must adopt cybersecurity best practices. There should be an understanding and trust that your privacy and data will be protected. Entities need to be held legally responsible for not implementing reasonable business practices.

My hope is that these ten rights will begin the much-needed and long-overdue conversation in Congress to guide a legislative solution that restores our privacy and protection online.

The American people can no longer wait while their data is being collected, shared and stolen on the web. The Internet can be a tool for more freedom and prosperity, but only if proper rules and guidelines exist. Our constituents tasked us to make those rules. It is now up to Congress to answer that call and bring our laws into the 21st Century.

Description: **super long article**

Link: <https://www.theguardian.com/environment/2018/oct/16/frankenbees-genetically-modified-pollinators-danger-of-building-a-better-bee>

Number of words: 4722

**Separated Summaries’ summary:** (it is separated into 11 sub-files, each contains 10 paragraphs of the original news article. Below is the summary of the summaries of those 11 sub-files):

in late april , dismayed beekeepers in germany looked on as whole colonies perished .

the bees were 10,000 times more potent than ddt .

technology for creating gm honeybees is in its infancy .

france , the netherlands and italy reported big losses .

millions of bees died .

**Direct summary:**

in late april , early may , dismayed beekeepers in germany looked on as whole colonies perished .

france , the netherlands and italy reported big losses .

the bees were 10,000 times more potent than ddt .

millions of bees died .

clothianidin developed by bayer crop science from a class of insecticides called neonicotinoids .

`` it was a disaster , '' german president says .

**Separated Summaries** (it is separated into 11 sub-files, each contains 10 paragraphs of the original news article. Below is the summaries of those 11 sub-files):

in late april , early may , dismayed beekeepers in germany looked on as whole colonies perished .

france , the netherlands and italy reported big losses .

the bees were 10,000 times more potent than ddt .

millions of bees died .

the technology for creating gm honeybees is in its infancy .

this is the plan to create a more resilient strain of honeybee .

if successful , it could lead to a hardier species , one that is resistant to natural and manmade hazards .

another , more controversial , response to the slump in bee populations .

jay evans heads the bee research lab at the us department of agriculture .

it is always hive-like - 30c and humid - in the narrow , windowless laboratory .

each disc contained 140 tiny plug holes in each of which a single honeybee embryo was growing .

evans calls them `` throw a lot of nature under the bus ''

these discs were passed to a third student at a separate workstation .

beye said , shaking his head in lingering disbelief .

in person , he is affable and professorial .

he says inferences of a bayer connection were totally false .

the purpose of their work was to understand the genetic basis for bee behaviour and health .

jay evans thinks his breakthrough gm bee should remain confined to the lab .

haefeker , a former tech entrepreneur , came to beekeeping late in life .

he , his wife and two sons returned home to germany in 2001 .

he developed an iphone app for breeders called iqueen and politics .

beekeepers fear genetic engineering of honeybees will introduce patents .

damage to the bee population is already threatening crops worldwide .

the un 's food and agriculture estimates that pollinators help farmers grow crops worth up to $ 577bn -lrb- ? 437bn -rrb- annually .

save the bees tops the list .

more than a dozen eu member states sought loopholes to stay the ban .

bayer officials say the industry is beginning to grow less reliant on chemicals .

despite the setback , haefeker remains defiant .

digital crop protection will feed the planet 's billions .

haefeker launched into what he called `` the big question ''

nobody , not even beekeepers , claimed an stake on the bees ' genetic code .

the free exchange of breeding materials from the queens and her eggs to the drones ' sperm has been encouraged to keep colonies genetically diverse .

apimondia is seeking to enshrine this freedom as a public good that nobody can own outright .

losses have doubled in the last 10 years .

he blames a host of factors , with disease and parasites such as the varroa mite chief among them .

he says he is closely watching what happens next in europe .

haefeker and arno bruder run their beekeeping enterprise on a field bordering two organic farms .

`` we need to keep an eye on this craziness , '' his text read .

'' .