

Due dates: unchanged
October 10 and 12 2023

Goal: Develop a context free feature grammar. Annotate different features in CoNLL format (including POS, NE, Ent, MeasEnt, ...)

Overview: A parse tree is a representation of syntactic structure, on which semantic interpretation can be based. You are to write a context free and a feature grammar that covers as much different text, as you can manage, but it must cover the validation data provided below. You may draw inspiration from anywhere, starting from the different CFG grammars in NLTK. You are to annotate text with POS, NE, Ent, MeasEnt features in CoNLL format.

Description:

- run your previously developed preprocessing pipeline (or an improvement) including a complete named entity module (NLTK is ok)
- develop a context free grammar to cover the validation data (additional coverage is always recognized)
- run the Earley parser in NLTK using your grammar
- ~~develop a context free feature grammar to cover the validation data (additional coverage is always recognized)~~
- ~~run the Earley's feature parser in NLTK using your grammar~~
- proofread the results, correct if necessary. Comment on strong and weak points in your report
- submit ~~both~~ grammars and your test data on October 10
- On October 11th, I will post a short challenge text. Run your pipelines and grammars over this challenge text and submit on October 12

CoNLL format

CoNLL format is a tabular format, where each word of a text begins a new line and each feature heads a new column, see CoNLL 2003 paper: <https://aclanthology.org/W03-0419.pdf>

- NLTK offers support to read and output CoNLL format. Description of CoNLL support for NLTK: <https://www.nltk.org/api/nltk.corpus.reader.conll.html>
- BIO and related schemes for span annotation in CoNLL format: <https://medium.com/@rongqianhui/named-entity-recognition-annotation-schemes-e684f9cd5a56>

Validation Data Run your preprocessing pipeline successively over the following sentences. The patterns of (1) have to be addressed and are sufficient for a B grade. Some patterns in (2 - 5) and other patterns from the inspirational (or other) data are optional. Your work should not be limited to patterns in the validation data. Describe the coverage of the grammar in detail in the report and motivate, why you prioritized the patterns you did.

1. *Two days ago, on Monday, John, who was sick for 5 days until Sunday, finally ate <a/5/all/the> red apple<s> from the fridge at the table in his office.*

Build up your grammar for (1) from shorter sentences, such as:

- (a) *John ate the apple at the table.*
 - (b) *John, who ate the apple from the fridge, was sick Wednesday.*
 - (c) *On Monday, John ate the apple in his office.*
 - (d) *Last week, on Monday, John finally took the apple from the fridge to his office.*
2. *Heart rate greater than 90 beats/minute*
 3. *Body temperature over 38 or under 36 degrees Celsius.*
 4. *Respiratory rate greater than 20 breaths/minute or partial pressure of CO2 less than 32 mmHg*
 5. *Leukocyte count greater than 12000 or less than 4000 /microliters or over 10% immature forms or bands.*

Interface Make sure that you have a way to input a new text that is then processed by the entire pipeline and displays all the annotations obtained on the screen without having to change the code. This can be as simple as a command line call and pretty print output for the parse trees. Provide an option to save/print. This is helpful for your own development, but essential for testing.

Deliverables and marking scheme:

On October 10 submit in Moodle:

- (2pts) the CFG you developed (Attrib 1, 12)
- ~~(2pts) the feature grammar you developed. Submit data for testing that you used (Attrib. 1, 12)~~
- (1.5pt) Parse trees and CoNNL output for CFG on development testing (Attrib. 1, 5)
- ~~(1.5pt) Parse trees and CoNNL output for feature grammar output on development testing (Attrib. 1, 5)~~
- ^{1pt}~~(2pts)~~ report (Attrib. 6) and demo files (Attrib. 5)

On October 12 submit in Moodle:

- ^{0.5pts}~~(1pts)~~ parse trees for run on challenge text (Attrib. 6)

Project now worth 5pts total.

The remaining 5pts will be rolled into the next project

For inspiration:

Systemic Inflammatory Response Syndrome

Rebanta K. Chakraborty, Bracken Burns.

In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan. 2023 May 29.

Excerpt

Systemic inflammatory response syndrome (SIRS) is an exaggerated defense response of the body to a noxious stressor (infection, trauma, surgery, acute inflammation, ischemia or reperfusion, or malignancy, to name a few) to localize and then eliminate the endogenous or exogenous source of the insult. It involves the release of acute-phase reactants, which are direct mediators of widespread autonomic, endocrine, hematological, and immunological alteration in the subject. Even though the purpose is defensive, the dysregulated cytokine storm can cause a massive inflammatory cascade leading to reversible or irreversible end-organ dysfunction and even death.

SIRS with a suspected source of infection is termed sepsis. Confirmation of infection with positive cultures is therefore not mandatory, at least in the early stages. Sepsis with one or more end-organ failures is called severe sepsis, and hemodynamic instability despite intravascular volume repletion is called septic shock. Together they represent a physiologic continuum with progressively worsening balance between pro and anti-inflammatory responses of the body.

The American College of Chest Physicians/Society of Critical Care Medicine-sponsored sepsis definitions consensus conference also identified the entity of multiple organ dysfunction syndrome (MODS) as the presence of altered organ function in acutely ill septic patients such that homeostasis is not maintainable without intervention.

Objectively, SIRS is defined by the satisfaction of any two of the criteria below:

Body temperature over 38 or under 36 degrees Celsius.

Heart rate greater than 90 beats/minute

Respiratory rate greater than 20 breaths/minute or partial pressure of CO₂ less than 32 mmHg

Leukocyte count greater than 12000 or less than 4000 /microliters or over 10% immature forms or bands.

In the pediatric population, the definition is modified to a mandatory requirement of abnormal leukocyte count or temperature to establish the diagnosis, as abnormal heart rate and respiratory rates are more common in children.

To summarize, almost all septic patients have SIRS, but not all SIRS patients are septic. Kaukonen et al. explained exceptions to this theory by suggesting that there are subgroups of hospitalized patients, particularly at extremes of age, who do not meet the criteria for SIRS on presentation but progress to severe infection and multiple organ dysfunction and death. Establishing laboratory indices to identify such subgroups of patients and the clinical criteria that we currently rely upon has been gaining prominence over recent years.

Several scores exist to assess the severity of organ system damage. The Acute Physiology and Chronic Health Evaluation (APACHE) score version II and III, Multiple organ dysfunction (MOD) score, sequential organ failure assessment (SOFA), and logistic organ dysfunction (LOD) score are to name a few.

History

With the advent of new concepts in pathophysiology and therapeutic interventions for sepsis in

the early 90s, there was an increasing need to identify a homogenous group of potential subjects for clinical trials investigating new innovative therapeutic strategies. Borne out of the plethora of emerging studies, one opinion was unanimous. An early, time-sensitive approach to diagnosis and intervention is necessary to impact patient survival and morbidity significantly. Identifying the subjects at any setting with easy-to-use standardized parameters, therefore, held the key. The American College of Chest Physicians/Society of Critical Care Medicine-sponsored sepsis definitions consensus conference held in Chicago, Illinois in August 1991 aimed to establish a standard group of clinical parameters to identify those subjects in any clinical setting easily. Thus was born the SIRS definition.

It underwent further modification in the second chapter of the meeting in 2001 in Washington, DC. This conference proposed a conceptual framework of the staging of sepsis using the PIRO acronym (predisposition, insult or infection, response, and organ dysfunction).

The goal of the initial definition was to be highly sensitive using easily available parameters across all healthcare settings. An unavoidable corollary of such a definition was, therefore, the lack of specificity. A few more relevant pitfalls of the SIRS definition, as has been pointed out in the literature, include the following:

- The universal prevalence of the parameters in an ICU setting

- Lack of ability to distinguish between beneficial host response from pathologic host response that contributes to organ dysfunction

- Distinguishing between infectious and non-infectious etiology purely based on the definition

- Lack of weight to each criterion ? e.g., fever and elevated respiratory rate have precisely the same significance as leukocytosis or tachycardia by the SIRS definition.

- Inability to predict organ dysfunction.

Kaukonen et al., in their study of over 130000 septic patients, established that one out of eight patients in their observational study of sepsis did not have two or more SIRS criteria. They also established that each criterion in the SIRS definition does not translate to an equivalent risk of organ dysfunction or death.

In the wake of this debate, in 2016, the European Society of Intensive Care Medicine and the Society of Critical Care Medicine (SCCM) created a task force that proposed Sepsis-3, a new definition for sepsis. The new definition excluded the establishment of SIRS criteria to define sepsis and made it more nonspecific as any life-threatening organ dysfunction caused by the dysregulated host response to infection. The task force claimed that sequential organ failure assessment (SOFA) has a better predictive validity for sepsis than SIRS criteria. It has better prognostic accuracy and the ability to predict in-hospital mortality. To reduce the complexity of calculating the SOFA, they introduced q SOFA.

Q SOFA

- 3 component assessment system with:
- Systolic blood pressure below 100 mm Hg
- Highest respiratory rate exceeding 21
- Lowest Glasgow coma score is under 15

Although the validity of q SOFA is limited in an ICU setting, it has consistently outperformed SIRS criteria in predicting organ dysfunction in a non-ICU and ER setting. The use of vasopressors, mechanical ventilation, and aggressive therapeutic interventions in ICU limit the efficacy of q SOFA.

Interestingly Hague et al., in their study of the utility of SIRS criteria in gastrointestinal surgery, patients also found it a useful criterion to identify postoperative complications.

, ,
”

Montreal Gazette:

Following the paper trail

A mill 30 kilometres east of Ottawa is our first stop on a quest to explore the environmental impact of producing a newspaper that’s read by 290,000 Montrealers every weekday (345,000 on Saturdays).

Author of the article: Marian Scott

Published Apr 20, 2007 ? Last updated Oct 28, 2009 ? 8 minute read

Eighty per cent of the newsprint used by The Gazette is made from wood chips at Papier Masson. Here, at the end of the line: rolls of paper weighing hundreds of kilograms. Eighty per cent of the newsprint used by The Gazette is made from wood chips at Papier Masson. Here, at the end of the line: rolls of paper weighing hundreds of kilograms. PHOTO BY GORDON BECK /The Gazette Article content “SUSHI,” says a handpainted sign on a patates-frites stand near the Papier Masson mill, the main employer in this town of 10,000.

Like the casse-croute’s menu, the 77-year-old newsprint mill on the banks of the du Lievre River has changed with the times.—

“People think pulp and paper is an old industry but it’s very high-tech,” says pulp production manager Sylvain Bussiere as photographer Gordon Beck and I don hard hats and steel-toed rubbers.

Article content

The mill 30 kilometres east of Ottawa is our first stop on a quest to explore the environmental impact of producing a newspaper that’s read by 290,000 Montrealers every weekday (345,000 on Saturdays) — and costs the equivalent of 186,816 trees a year.

Papier Masson supplies 80 per cent of the newsprint, made from wood chips, on which The Gazette is printed. The other 20 per cent comes from Atlantic Paper, a producer of 100-per-cent recycled newsprint in Toronto.

Wendy Desmarteaux, The Gazette’s vicepresident of operations, says Papier Masson is a preferred supplier because of p“rice and the newsprint’s excellent runnability — uniform texture and ability to hold ink.

It requires much less energy to produce recycled newsprint but costs the same because of the high cost of waste paper and the expense of removing glass and plastic from it, according to Gerry Murray, Atlantic’s director of mills.

“Plastic bottles are our nemesis,” Murray says.

Papier Masson’s customers also include the Ottawa Citizen, National Post, USA Today, Denver Post and other papers as far afield as Egypt and Brazil.

In 2000, the mill, now owned by White Birch Paper of Greenwich, Conn., invested \$158 million in a thermomechanical pulper that churns out 760 metric tonnes of wood pulp every day.

It feeds a paper machine that produces 665 tonnes of newsprint daily — enough to print two

weeks' worth of The Gazette.

The investment transformed an old-fashioned, foul-smelling mill into one of the most efficient in Canada, says Bussiere, an Ecole Polytechnique-trained mechanical engineer.

Overhead, the landmark twin water towers that used to store chemicals for pulp are now empty. That's because the new process uses heat to soften the wood chips and powerful metal disks to grind them into pulp, so the mill smells pleasantly of sawdust.

The new pulp refiner has cut greenhousegas emissions by 54 per cent, water use by 34 per cent and organic waste by 40 per cent, earning the mill a first-place Eco-geste award from the Quebec Environment Department in 2003.

It also eliminated 55 per cent of the jobs: 221 work here compared with 485 seven years ago.

The plant now recycles 98 per cent of its solid waste; wood scraps are sent out to make particle board and organic waste is composted and trucked out to farmers' fields. All liquid effluent is treated to remove pollutants before it is returned to the du Lievre River, which flows into the Ottawa River 1.5 kilometres downstream. "It's crazy, until 1985 we dumped everything back in the river," says Bussiere, whose first ambition was to work in the environmental field.

But the massive refiner's powerful, 36megawatt motor — "like a Boeing," Bussiere notes — consumes enormous amounts of energy — and wood.

When the James MacLaren Company, a local lumber dynasty, opened the mill in 1930, the abundant forest that once grew here kept it supplied with pulp logs.

"The du Lievre River runs through a pine region, and vast forests of spruce are waiting for the woodsman's axe and the pulp mill and paper mill," the Toronto Globe reported in 1895. "There is sufficient spruce standing on the banks of this river to supply the largest paper mill in the world at the present time for a hundred years, so that it might be truthfully stated that the supply is inexhaustible."

But the inexhaustible forest was logged out years ago.

Nowadays, wood chips — the mill's raw material — come from the Abitibi and northern Laurentians, a trip of up to seven hours by truck.

Out in the yard, a 66-foot semi-trailer — one of 240 that arrive every week, laden with chips — is preparing to dump its 20-tonne load. The rig drives onto a mechanized ramp that, like some robotic Godzilla, rises into the air, raising the massive truck almost perpendicular to the ground until its contents spill out. A conveyor belt carries the wood chips to the pulper.

The chips are a sawmill by-product made from remnants too small for lumber, Bussiere explains.

Up to 85 per cent of the fibre used to make pulp in Canada comes from wood chips, says Jean-Pierre Martel, senior vice-president of sustainability at the Forest Products Association of Canada. "In the 1960s and '70s, it was all pulp logs."

Martel said the forestry industry has made great strides in making logging more sustainable.

In Canada, 94 per cent of forest is publicly owned and 85 per cent is certified under one of three systems that ensure logging companies take preservation of natural ecosystems into account. This includes reducing impact on waterways and preserving wildlife habitat.

"Less than seven per cent of the world forest is third-party certified," Martel noted.

But Tracy London, newspaper campaigner for Markets Initiative, says the vast volume of wood used to feed the pulp and paper industry is cause for concern.

Canada is the world's largest producer and exporter of newsprint and wood pulp, producing 7.8 million tonnes of newsprint and 25.2 million tonnes of pulp and 6.7 million tonnes of printing and writing paper in 2005, according to the federal government report, The State of Canada's Forests 2005-06.

Jim Fyles, McGill University professor and scientific director of the Sustainable Forest Management Network, said awareness of the importance of maintaining forest ecosystems has improved in the forestry industry.

“There is a good understanding in order to have a sustainable forest you need to replace the trees that are cut.”

But Fyles says scientists are uncertain about how much harvesting of the forest should be allowed. “We need to be able to sustain the forest because it sustains a lot of people.”

Inside the mill, the paper machine is spinning out 1,300 metres of newsprint a minute. Circular saws slice the 8.3-metrewide roll into custom widths ordered by customers.

The rolls coming off the paper-winder now are for the Chicago Sun-Times and Philadelphia Newspapers Inc., which publishes the Philadelphia Inquirer and Daily News. Someday soon, readers in Philly and Chicago will be poring over the now pristine paper gliding on a conveyor belt to the shipping area.

Pressman Pierre Belanger joined The Gazette when Expo 67 was still in the planning and Mother Martin’s was the watering hole of choice for thirsty Gazette staffers.

Forty years and two printing presses later, Belanger, 60, figures the 2002 move to the newspaper’s purpose-built printing plant on St. Jacques St. W. in Notre Dame de Grace was his last.

“Once you have a taste of that, you don’t want to go back,” he says of the two computer-controlled offset presses that now print The Gazette.

Technological upgrades have also saved costs and reduced environmental impacts, says Desmarteaux.

Soy-based inks have replaced petroleum products The Gazette used in the 1960 vintage letter press at its former location in Old Montreal. That has significantly reduced emissions of volatile organic compounds (VOC), once a common by-product of the printing industry.

Vegetable inks also reduce toxic waste when newspapers are recycled.

And computerization recently eliminated the need for film, and the chemicals used to process it, in the printing process.

In the past two decades, The Gazette has cut newsprint consumption almost in half — 17,000 tonnes last year, compared to 31,000 tonnes in 1987.

The savings are due to fewer pages and a switch to a smaller broadsheet.

Desmarteaux estimates the paper will use 15,000 tonnes of newsprint this year.

“This is three weeks of inventory,” she says, indicating the humidity-controlled storage area where gigantic rolls of newsprint are stacked six high.

The newer presses have also eliminated waste at the start of each press run — 1,500 copies are scrapped compared to three times that on the old press, Desmarteaux adds.

After printing — a complex ballet where colour and black-and-white pages somehow sort themselves into their proper sections — the newspaper is cut, folded, bundled and transported by conveyor belt to the loading dock.

Delivery: adults in cars

While the city sleeps, Mike Raffoul checks his watch. “They should have started at 2,” he frets.

It’s 2:30 a.m. and the presses at The Gazette’s St. Jacques St. plant still haven’t started printing the Saturday sports final — the last edition of the paper, which goes to homes and stores in central Montreal.

Raffoul is an independent contractor responsible for making sure 40,000 subscribers from St. Leonard to Lachine get their Gazettes on time.

He's among 13 distributors who employ 60 subcontractors and 800 carriers to deliver the paper. The distribution contractors' 80 trucks and the carriers' 800 cars travel a total of about 22,000 kilometres every day.

If the paper comes off the press late, Raffoul and his army of 500 carriers will have to make up the time.

On top of that, it's raining and this morning's paper is thicker than usual because of a special advertising supplement.

"All we need is a good snowstorm and we're all set," Raffoul quips.

Twenty-five years ago, kids on foot delivered most Gazettes and other newspapers across North America.

District managers driving Gazette vans dropped off the bundles at carrier's homes or at drop points.

But since the early 1990s, The Gazette has been delivered by adults in cars and today's routes are five times bigger — 200 to 300 homes — than they were 25 years ago.

Suburban sprawl, the death of afternoon papers, contracting-out of newspaper distribution and parents who don't let kids go out unsupervised have all contributed to the change, Raffoul says.

He supplies the plastic bags to protect newspapers in wet weather. Carriers pay 1.5 cents per bag. Raffoul knows biodegradable bags would be better for the environment but doesn't want to increase carriers' costs.

At the Rose Bowl parking lot at St. Jacques St. and Cavendish Blvd., 25 carriers are already waiting in their idling cars for bundles.

Among them is Gabriel, a 37-year-old father of two who came to Canada from Romania three years ago. He also happens to be my paper carrier. Gabriel is one of the few carriers who actually walks his route.

He starts at 4 a.m., finishes by 6, then drives his kids to school and goes to his day job as an accounts-receivable clerk.

"It's like jogging," says Gabriel, who picks up the bundles in his 2005 Corolla, then parks and delivers the papers on foot.

He was thinking of quitting the route recently but decided he actually enjoys the pre-dawn workout.

"I realized it's not just for the money, it's for me."