Lusha

Email Signature

Information Extraction

Project

Develop an efficient and accurate pipeline to extract key information from email signatures using Named Entity Recognition (NER).

☑ Why? To enhance the company's core value by unlocking a new source of contacts and streamlining the extraction process.

☑ Added Value of NER: Improve information quality by removing unnecessary data and automating the extraction process (e.g: now:" hannah welch (she, her)", NER: "hannah welch").

Journey

___

1. **Start with Base BERT/Roberta**

   └────> 2. Improve and Customize

2. **Improve and Customize**

   ├────> 2a. Create synthetic dataset

   ├────> 2b. Finetune model

   └────> 2c. Perfect results

3. **Recognize need for labeled dataset**

   ├────> 3a. Labeling true signatures hard/costly

   └────> 3b. Generate signatures with GPT-4

4. **Model Fine-tuning**

   └────> 4a. Adjust model with new dataset

Pipeline Overview

1 Data Preprocessing for Templates

2 Generation of Templates and Signatures

3 Model Training

4 Inference

Template Generation

Programmatically (training set)

Warm regards,
Mr. {name},
{job_title}
{location}
{email} | Call: {phone}
{webpage}/vrpdwnrlmoumqotwmwbsmlquvtgyw.com
linkedin.com/in/{social_networks}

GPT-4 API (validation and test set)

Sincerely,
ryan rodriguez, pg,
wellsite geologist
adams inc | geology & exploration division
0769 jacob pass suite 569
allenville, in 21777
office : (555) 123 - 9876
r.rodriguez@example.com
linkedin : linkedin.com/in/ryanrodriguez
* please note : this email may contain confidential information. if you are not the intended recipient, please delete this message, and any attachments, without copying, printing or forwarding.

🐦 Lusha

What is seqeval?

Python library for sequence labeling models (e.g., NER)

Considers entire label sequences for accurate assessment

Accounts for context and continuity

Calculates precision, recall, and F1-score

Comparables and benchmarks

	Size	Duration	F1-score
BERT fine-tuned	1,000	4min	0.642
	10,000	12min	0.763
	100,000	2h	0.723

		Duration	F1-score
RoBERTA fine-tuned		4min	0.83
		16min	**0.888**
		2h	0.871

dslim/bert-base-NER-uncased   0.081

Lusha

GPT-4 generated

signature with real tags

warm regards, daniel berg ii, mrics
surveyor, planning and development
lane - matthews
022 miller mountain suite 325
south derek, az 98966
phone : (792) 901 1502
email : daniel.berg@lane-matthews.com

warm regards, daniel berg ii, mrics
surveyor, planning and development
lane - matthews
022 miller mountain suite 325
south derek, az 98966
phone : (792) 901 1502
email : daniel.berg@lane-matthews.com

RoBERTa
fine-tuned

VS

dslim/bert-base-

NER-uncased

warm regards, daniel berg ii, mrics
surveyor, planning and development
lane - matthews
022 miller mountain suite 325
south derek, az 98966
phone : (792) 901 1502
email : daniel.berg@lane-matthews.com

GPT-4

take care, traci lopez, m.ed.
further education lecturer
white, wilson and riggs 16000 lara islands
east jessica, il 68548 office : (555) 234 - 5678
mobile : (555) 345 - 6789
email : traci.lopez@wwr.edu linkedin :
linkedin.com/in/tracilopez
twitter : @tracilopez_edu confidentiality disclaimer : this
email may contain confidential information for the
addressee. unauthorized use or disclosure is
prohibited.

---

unbeatabledealsonlaptopsanddesktops!
upgradeyourtechgametoday!
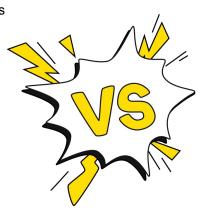shophere:[insertfakeurl]

---

kind regards,

 jonathan conley, m.a.ed.
special educational needs teacher
bishop group 978 gillespie fork apt. 526
ibarraborough, nh 22836 office : (555) 315 -
2468
mobile : (555) 315 - 2469
email : jonathan.conley@bishopgroupschools.org
inspiring minds, building futures.

RoBERTa fine-tuned

take care, traci lopez, m.ed.
further education lecturer
white, wilson and riggs 16000 lara islands
east jessica, il 68548 office : (555) 234 - 5678
mobile : (555) 345 - 6789
email : traci.lopez@wwr.edu linkedin :
linkedin.com/in/tracilopez
twitter : @tracilopez_edu confidentiality disclaimer : this
email may contain confidential information for the
addressee. unauthorized use or disclosure is
prohibited.

---

unbeatabledealsonlaptopsanddesktops!
upgradeyourtechgametoday!
shophere:[insertfakeurl]

---

kind regards,

 jonathan conley, m.a.ed.
special educational needs teacher
bishop group 978 gillespie fork apt. 526
ibarraborough, nh 22836 office : (555) 315 -
2468
mobile : (555) 315 - 2469
email : jonathan.conley@bishopgroupschools.org
inspiring minds, building futures.

Inference on real signatures

hannah welch (she, her)
senior partner sales manager
tableau
07398724561
hwelch@salesforce.com
(mailto:hwelch@salesforce.com)
please note a change in my working hours to
tues-friday.

percy brock - director
outland group ltd.
m: 44 (0) 7848 455048 w:
www.outlandgroupltd.com

t +32 3 800 76 49 | m +32 490 66 96 28 baarbeek
1
2070 zwijndrecht belgium
stephensonweg 6 4207 hb gorinchem nederland
michael.vantveer@agidens.com
www.agidens.com

ubt (eu) ltd
exchange place,
poseidon way warwick, warwickshire cv34 6by
united kingdom
add us to your address book
(https://emea01.safelinks.protection.outlook.com/?url=https%3a%2f%2fubteam.us12.list-manage.com%2fvcard%3fu%3d409dcd49fa1a58bbf468b72
2c%26id%3d42a2bee134&data=04%7c01%7c%7cecbb0dabe2ee49cf061508d98e6f919d%7c84df9e7fe9f640afb435aaaaaaaaaaaa%7c1%7c0%7c6
37697432119751235%7cunknown%7ctwfpbgzsb3d8eyjwijoimc4wljawmdailcjqijoiv2lumziilcjbtii6ik1hawwilcjxvci6mn0%3d%7c1000&sdata=ob5ynekw
vijfl9maaq5cnfbovrl9yqinx0x4aujh5yc%3d&reserved=0)
registered in england and wales no. 4938684.
want to change how you receive these emails?
you can
update your preferences or
(https://emea01.safelinks.protection.outlook.com/?url=https%3a%2f%2fubteam.us12.list-manage.com%2fprofile%3fu%3d409dcd49fa1a58bbf468b72
2c%26id%3d42a2bee134%26e%3dabd362ac6f%26c%3d37c11146eb&data=04%7c01%7c%7cecbb0dabe2ee49cf061508d98e6f919d%7c84df9e7fe
9f640afb435aaaaaaaaaaaa%7c1%7c0%7c637697432119751235%7cunknown%

Next Steps

1 Refine Templates

2 Address Leakage

3 Improve Performance on Low-quality Signatures

4 Investment Decision

- Template refinement, or
- preprocessing for better quality signatures.

Lusha

Thank you!

Lusha