FROM TEXT TO KNOWLEDGE,
FROM KNOWLEDGE TO DECISIONS
A technology (AI-NLP) and Prevyo, a product line

10 peoples (PhD students, engineers, research engineers)
13 papers since 2018, 2 awards
Patents pending
Email: Simple & massively used

330 billion e-mails per day worldwide
(source: Statista, 2022)
# The email dilemma

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Massive</td>
</tr>
<tr>
<td>Universal</td>
<td>Disturbing</td>
</tr>
</tbody>
</table>
Data you can find in emails

Structured

- Sender
- Recipient(s)
- Date
- Priority

Unstructured

- Object
- Body
- Attachments
Hi Philippe,

I am Penbase sales manager and I am contacting you because we would like to know more about your Prevyo solution. To give you an idea, we are a 40 people company and all of us spend an average 1 hour per day managing emails.

Could we arrange a meeting next week in your offices?

Thank you very much,
Nicolas Gagean

Nicolas Gagean - Directeur commercial
nicolas.gagean@penbase.com
41 30 78 12 86
Penbase - www.penbase.com
Espace Bocaud - 42 rue de la Pierre Plantée
34830 Jacou - France
## Two kinds of information

<table>
<thead>
<tr>
<th>Facts</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Your invoice is available”</td>
<td>“Please call me back at 4pm”</td>
</tr>
<tr>
<td>“I studied your project”</td>
<td>“Can you send the agreement?”</td>
</tr>
</tbody>
</table>
• Identification of tasks in e-mails since the 90’s (Khosravi et Wilks, 1999)
• Lot of research thanks to the Enron corpus (Klimt et Yang, 2004).

Dialogue act classification

● Statistical approaches based on n-grams (Carvalho et Cohen, 2006)
● Supervised learning (Khoo et Albrecht, 2006 ; Lampert et al. 2009 ; Kalitvianski 2018)
● Unsupervised learning (Joty et al. 2011 ; Lampert et al. 2014)
● Semi-supervised learning (Jeong et al. 2009)
● Rule-based system (Scerri et al. 2010)
An approach based on semantic role labeling

No email dataset is available for French

Hybrid approach (Mekaoui et al. 2020)
- Statistical (machine learning > transfer learning)
- Symbolic (rules, ontologies)
Collaborative research projects in progress…

- **POPCORN**, 36 months, funded by AID: industrial research for semi-automated enrichment of a knowledge base for security and defense customers.

- **XP-Event**, 36 months, funded by Occitanie: Event detection and prediction in emails.
Demo available online:
https://pss.prevyo.com
Thank you for your time!

cedric.lopez@emvista.com