When is my flight?

Semantic data extraction in KMail and Nextcloud Mail

FOSDEM 2024

Volker Krause

vkrause@kde.org
@VolkerKrause@floss.social
Motivation

- Hard to find relevant information in ad overloaded HTML email
- Tedious to manually enter all booking information into calendar/travel app/etc.
- Had to figure out my departure gate in an unreadable email, looked at source code, things got out of hand...
api_key: "your_api_key"
Reality

- JSON, Microdata or “creative” JSON
- Incomplete or incorrect data
  - Wrong timezones, mismatching address and coordinates, etc
- Limited use
  - Some airlines, few hotels, practically no train operators
  - Less common in Europe than in the US, very rare in Asia (?)
Other sources

• Structured
  – Barcodes, in particular for international flights/trains

• Semi-structured
  – Apple Wallet passes

• Unstructured
  – Generic extractors for e.g. boarding passes
  – 130+ provides specific scripts
Post-processing

- Normalization
- Apply context and domain knowledge to resolve ambiguities
- Augment from other data sources
  - Airport/train station code lookup in Wikidata/OSM
  - Timezone from country/location
- Validation and plausibility checks
Buchungsbestätigung Deutsche Bahn (Auftrag: [对此内容进行审核])

| From: "buchungsbestaetigung@bahn.de" <buchungsbestaetigung@bahn.de> |
| To: |
| Date: |
| Spam Status: Spamassassin |

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>02/02/2024 10:57 CET</td>
<td>02/02/2024 15:15</td>
</tr>
<tr>
<td>ICE 650</td>
<td></td>
</tr>
<tr>
<td>Class: 2 Coach: 4 Seat: 24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>02/02/2024 15:40</td>
<td>02/02/2024 17:35 CET</td>
</tr>
<tr>
<td>ICE 14</td>
<td></td>
</tr>
<tr>
<td>Class: 2 Coach: 23 Seat: 28</td>
<td></td>
</tr>
</tbody>
</table>

[Links: Show Calendar, Add to Calendar, Add to KDE Itinerary, Send to]
Integration in Nextcloud Mail

Buchungsbestätigung (Auftrag BPM83K)
Jos Poortvliet to christine.schott.nc@gmail.com

Stuttgart Hbf
12/27/2019
12:00 AM →
Nürnberg Hbf
06/28/2019
4:18 PM

Nürnberg Hbf
06/28/2019
5:01 PM →
Berlin Hbf (tief)
12/27/2019
12:00 AM

Hi Christine,
Here's also a train ticket example - I went to Stuttgart, and back to Nuremberg.

Buchungsbestätigung (Auftrag BPM83K)
Sehr geehrter Herr Poortvliet,
Integration in Thunderbird

Image showing an email with flight information for Eurowings flights.

**Bordkarten für Ihren Eurowings-Flug der Buchung**

- **STR**
  - Mi., 19. Juni 2019, 09:40 MESZ
  - EW 2530
- **VLC**
  - EW 2531
- **VLC**
  - Mi., 26. Juni 2019, 12:30 MESZ
  - EW 2531
- **STR**
  - Mi., 26. Juni 2019, 14:50 MESZ

**Add to KDE Itinerary**

**Download**

---

**Ihre Bordkarten**
Integration in KDE Itinerary

My Itinerary

Friday, 2/2/24

KDE e.V. Office to Berlin Hbf
10:05 AM

ICE 650 → Köln Hbf
10:57 AM
10:57 AM Berlin Hbf PL 13
5 intermediate stops 3:15 PM Köln Hbf PL 7

10°C / 12°C

ICE 14 → Bruxelles Midi 3:40 PM
3:40 PM Köln Hbf PL 5 D-G
3 intermediate stops 5:35 PM Bruxelles Midi PL 10

Event

FOSDEM is a free event for software developers to meet, share ideas and collaborate. Every year, thousands of developers of free and open source software from all over the world gather at the event in Brussels.

Location
Université Libre Bruxelles
Brussels Belgium

Start Time
2/3/24 10:00 AM

End Time
2/4/24 5:00 PM

Website
https://fosdem.org/2024/

Documents and Tickets
No documents attached to this reservation.
Conclusion

- Semantic annotations in emails enable productivity features
- Incorrect data is worse than no data
- Please add semantic annotations to emails you produce!
- Donations of sample documents are very welcome :)

Thank you!

Questions?

Meet KDE: KDE stand, building H, level 1

Contact: Email: vkrause@kde.org Matrix: @vkrause:kde.org Mastodon: @VolkerKrause@floss.social
References

- Semantic data extraction engine: https://invent.kde.org/pim/kitinerary
- KMail plug-in: https://invent.kde.org/pim/kdepim-addons
- Nextcloud Mail integration: https://github.com/nextcloud/mail
- Thunderbird plug-in: https://invent.kde.org/broulik/itinerary-travelbird
- KDE Itinerary: https://kde.org/for/travelers and https://apps.kde.org/itinerary