



# 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...



# schema.org annotations

```
{
  "@context": "http://schema.org",
  "@type": "FlightReservation",
  "reservationNumber": "RXJ34P",
  "underName": { "@type": "Person", "name": "Eva Green" },
  "reservationFor": {
    "@type": "Flight",
    "flightNumber": "110",
    "airline": { "@type": "Airline", "name": "United", "iataCode": "UA" },
    "departureAirport": { "@type": "Airport", "name": "San Francisco Airport",
                          "iataCode": "SFO" },
    "departureTime": "2027-03-04T20:15:00-08:00",
    "arrivalAirport": { "@type": "Airport", "iataCode": "JFK",
                       "name": "John F. Kennedy International Airport" },
    "arrivalTime": "2027-03-05T06:30:00-05:00"
  }
}
```



## 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




# Integration in KMail

## Buchungsbestätigung Deutsche Bahn (Auftrag: )





From: "buchungsbestaetigung@bahn.de" <buchungsbestaetigung@bahn.de>  
To: [mailto:buchungsbestaetigung@bahn.de](mailto:mailto:buchungsbestaetigung@bahn.de)  
Date:   
Spam Status: **Spamassassin**

Berlin Hbf [13] 02/02/2024 10:57 CET ⊕ Class: 2 Coach: 4 Seat: 24	→ ICE 650	Köln Hbf [7] 02/02/2024 15:15
Köln Hbf [5 D-G] 02/02/2024 15:40 ⊕ Class: 2 Coach: 23 Seat: 28	→ ICE 14	Bruxelles Midi [10]  02/02/2024 17:35 CET

 Show Calendar

 Add to Calendar

 Add to KDE Itinerary

 Send to  [mlg@23](mailto:mlg@23)



# Integration in Nextcloud Mail

The screenshot displays the Nextcloud Mail interface. On the left, a sidebar shows folders like 'All inboxes', 'christine.schott.nc@gmail.com', 'Inbox', and 'Favorites'. The main area shows an email from 'Jos Poortvliet' with the subject 'Buchungsbestätigung (Auftrag BPM83K)'. The email content includes a train ticket confirmation for Stuttgart Hbf to Nürnberg Hbf and Nürnberg Hbf to Berlin Hbf (tief). A context menu is open over the ticket details, offering options to 'Import into Family', 'Import into Personal', and 'Import into Work'. The DB logo is visible at the bottom right of the email content.

**Buchungsbestätigung (Auftrag BPM83K)**

Jos Poortvliet to christine.schott.nc@gmail.com

<b>Stuttgart Hbf</b> 12/27/2019 12:00 AM	IC 2069 →	<b>Nürnberg Hbf</b> 06/28/2019 4:18 PM
<b>Nürnberg Hbf</b> 06/28/2019 5:01 PM	ICE 1002 →	<b>Berlin Hbf (tief)</b> 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

The screenshot shows the Mozilla Thunderbird interface displaying an email from noreply@eurowings.com. The email subject is "Bordkarten für Ihren Eurowings Flug der Buchung [REDACTED]" and it was received on 24.06.19 at 13:34. The email content displays a flight itinerary for two days:

<b>STR</b> Mi., 19. Juni 2019, 09:40 MESZ	 EW 2530	<b>VLC</b> Mi., 19. Juni 2019, 11:50 MESZ
<b>VLC</b> Mi., 26. Juni 2019, 12:30 MESZ	 EW 2531	<b>STR</b> Mi., 26. Juni 2019, 14:50 MESZ

At the bottom of the itinerary table, there are two buttons: "Add to KDE Itinerary" and "Download".

In the bottom right corner of the email preview, the "Eurowings" logo is visible.

**Ihre Bordkarten**




# Integration in KDE Itinerary

☰ My Itinerary ⋮

📅 Friday, 2/2/24

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

🚶 U3 S5 🚶 One change   
Click to extend

🚆 ICE 650 → Köln Hbf 10:57 AM

10:57 AM	<b>Berlin Hbf</b>	Pl. 13
▼ 5 intermediate stops		
3:15 PM	<b>Köln Hbf</b>	Pl. 7


🌤️ 10°C / 12°C

🚆 ICE 14 → Bruxelles Midi 3:40 PM

3:40 PM	<b>Köln Hbf</b> Germany	Pl. 5 D-G
▼ 3 intermediate stops		
5:35 PM	<b>Bruxelles Midi</b>	Pl. 10

☰ Itinerary   🔍 Plan Trip   📄 Passes & Programs

☰ < My Itinerary > Event

  
FOSDEM

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

📍 View Indoor Map   🗺️ View on Map   > Navigate   ⋮

**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.

☰ Itinerary   🔍 Plan Trip   📄 Passes & Programs



## 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](mailto:vkrause@kde.org)

Matrix: [@vkrause:kde.org](https://matrix.to/#/@vkrause:kde.org)

Mastodon: [@VolkerKrause@floss.social](https://floss.social/@VolkerKrause)



## 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>