

Data Engineer Assignment

Invoice suggestion based on historical data

Usually data is extracted from scanned invoices digitally. For a better user experience, we would like to have a RESTful web service that enhances the process and informs us about incorrect extraction based on historical data.

Although it's a basic service that exposes a few RESTful endpoints, we are keen to see your design decisions, well-structured code, and of course testability. Please make sure your solution is ready for production and similarly that future requirements can be added without too many structural changes.

Please also document the development lifecycle (e.g installation, running tests, ...) and package your solution as a deployable container.

This task is divided into two phases, and should take you no more than 4 hours.

Evaluation Criteria:

- Well-organized and testable modern Python code
- MongoDB should be the main database
- Test cases
- Production readiness

Notes:

- In the ingestion phase, the change ratio for document and contact is 30:1, meaning we expect invoices to be created 30 times faster than updating a contact.
- Authentication and authorization are **not required**.
- Please feel welcome to use a web framework of your choice.
- Please feel welcome to use a testing library of your choice.

1. Data Ingestion

The service should expose two rest endpoints:

- Creating a new invoice with contact information
- Updating one contact

Payload for the creating invoice with contact:

```
{
  "_id": "unique-invoice-id", // String (unique id is guaranteed)
  "organization": "organization-id", // String
  "createdAt": "2021-10-11T09:53:31.339Z", // Date
  "updatedAt": "2021-11-29T13:15:19.500Z", // Date
  "amount": {
    "currencyCode": "EUR", // String (Valid currency symbol)
    "value": 26.3 // Float (positive float number)
  },
  "contact": {
    "_id": "unique-contact-id", // String (unique id is guaranteed)
    "iban": "DE8810050001310032358", // String
    "name": "Contact Name", // String
    "organization": "organization-id" // String
  },
  "invoiceDate": "2021-10-11T00:00:00.000Z", // Date
  "invoiceId": "VR210230898" // String
}
```

Payload for updating a contact:

```
{
  "_id": "unique-contact-id", // String (unique id is guaranteed)
  "iban": "DE8810050001310032358", // String
  "name": "Contact Name", // String
  "organization": "organization-id" // String
}
```

2. Data Suggestion

The service should implement **at least one** of the following endpoints:

- Suggest the invoice contact based on previous invoices.

Request:

```
{
  "organization": "organization-id", // String
  "contactName": "partial input or extracted contact name" // String
}
```

Response:

```
{
  "suggestedContact": "suggested contact", // String
  "confidence": 0.8 // float (i.e 1 means 100% confidence)
}
```

- Indicate an abnormal amount for one invoice

Request:

```
{
  "organization": "organization-id", // String
  "contactName": "partial input or extracted contact name", // String
  "amount": 1000.0 // float
}
```

Response:

```
{
  "abnormal": true // boolean
}
```