1. Sizing: How many documents do you assume need to be indexed? e.g. <500.000?

It's hard to determine the exact number of documents to index. As we will clarify in the following questions, we are not really sure what sources will be indexed and how many "subcategories", e.g. courses, we want to crawl. This depends a lot on the Lufthansa preferences.

A short research gave me these numbers:

Udemy stats till May 2017

- They have 15 million+ students
- *20,000+ expert instructors*
- 140+ Topics or course categories
- Roughly 50,000 plus courses

As you can see, the size depends on the Lufthansa preferences and how much information they want us to index/crawl.

2. How often should the contents be updated.

We think at this point we are not able to answer this question exactly. Although we think a daily crawl would be sufficient, we think it depends on the type of source.

3. Preprocessing and data cleansing: Which MOOCs are interesting for Lufthansa (areas of interest).

At the moment we have not really specified an explicit list. All we have is a list provided by Udo Sonne. I think the Lufthansa management / Udo Sonne can specify the most relevant of these.

 https://docs.google.com/spreadsheets/d/1x40RjzOl6NNB26M7CQ4AIDgkWI9X2 BB-mWYbeGo2XWU/edit?usp=sharing

We focused so far on the more popular ones like https://udemy.com and https://udemy.com.

4. Did you already create a sitemap for those main entry points (based on the standard

at sitemaps.org) OR any other crawling pattern?

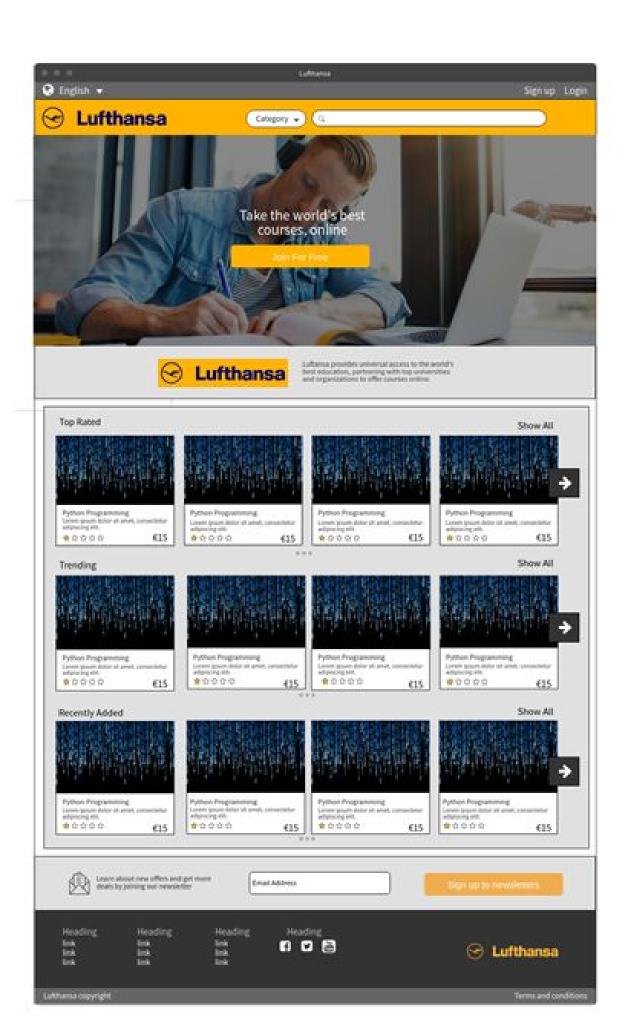
So far we concentrated on the API information. We analyzed some of the given json files and discussed the structure of these as we didn't plan to crawl the sites directly.

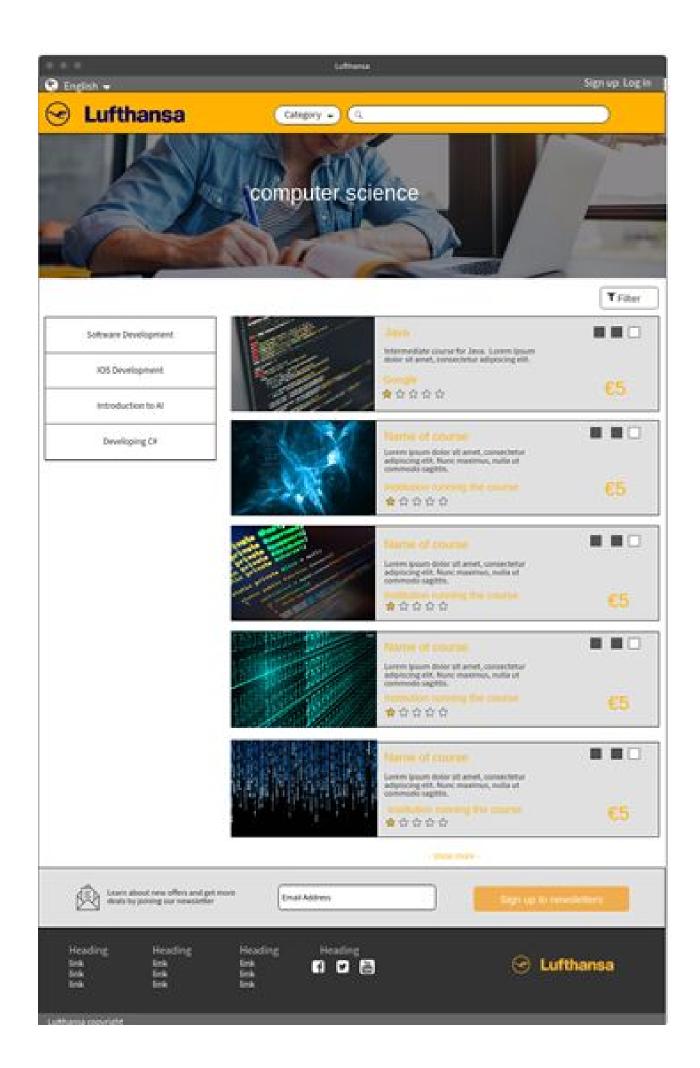
Of course we could create a sitemap (points of entry) in xml for you. For a better understanding of our previous approaches see point 8.

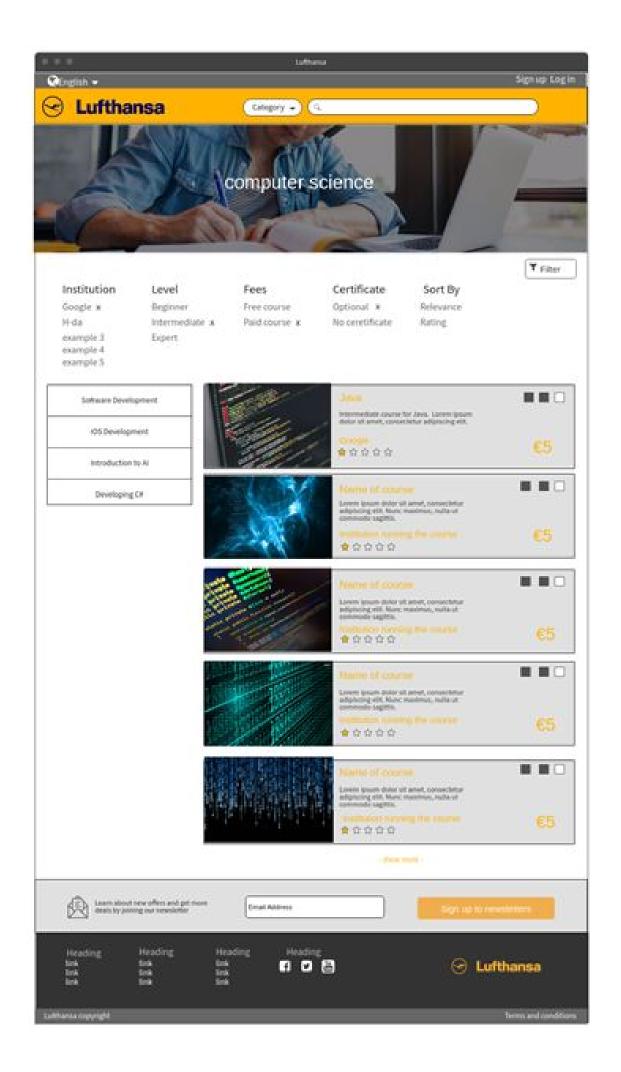
5. UI mockups: What should the search app look like?

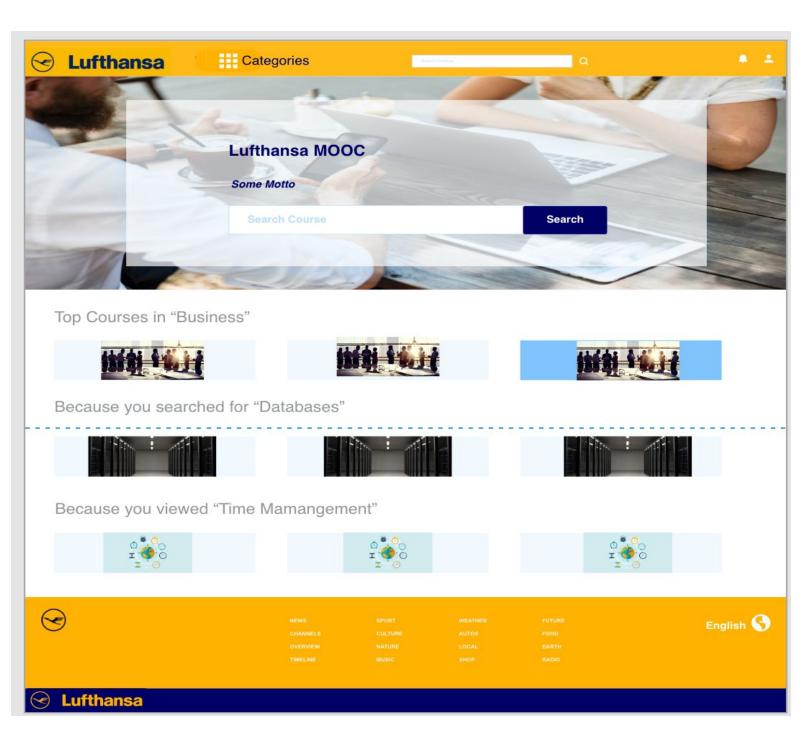
Concerning the customization, we have a few UI mockups build by the design team. They are at a very early state and we'll continue the development as soon as we know the technical possibilities.

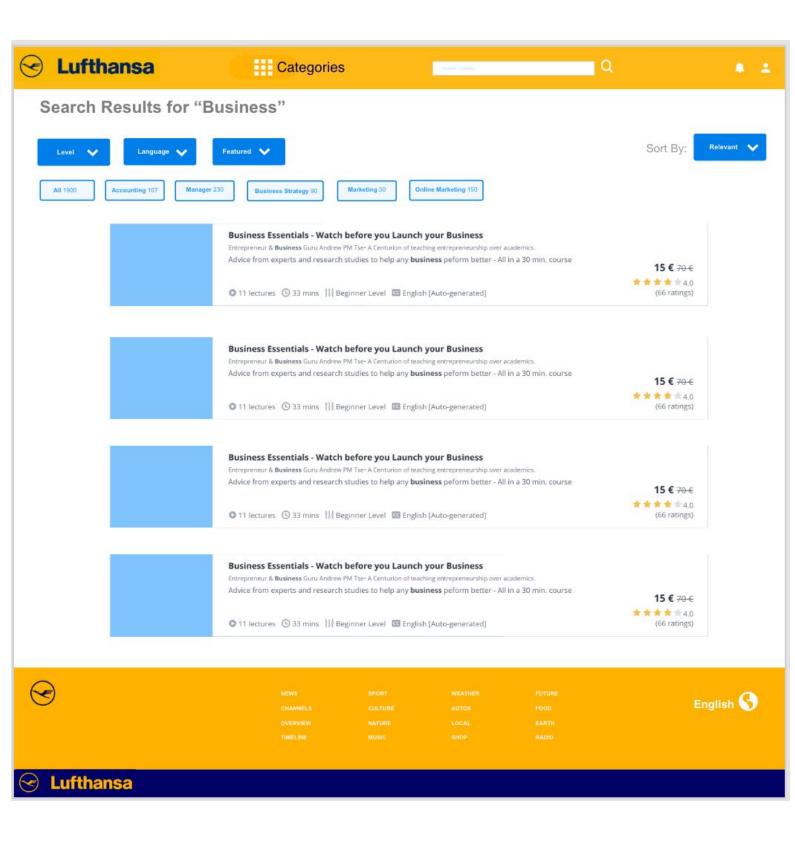
See the following pages..











Business Essentials

What Will I Learn:

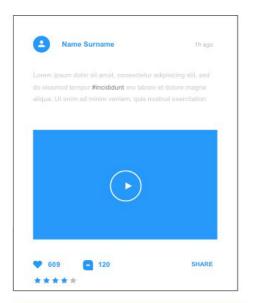
- Have a basic understanding of the merits of going into business and having an online business.
- Have a basic understanding of the merits of going into business and having an online business
- Have a basic understanding of the merits of going into business and having an online business.
- Have a basic understanding of the merits of going into business

Requirements

- -No requirements for this course other than the desire to be a success in business
- -Elite Percent courses are set out in a concise, structured format giving you synthesised learning across the major areas of your life or ones you deem most important to your immediate goals.

Description

Elite Percent courses are set out in a concise, structured format giving you synthesised learning across the major areas of your life or ones you



Start Course

Compare to Other Business Fundamentals Courses

Business Essentials - Watch before you Launch your Business Advice from experts and research studies to help any business perform better - All in a 30 min. course 15 € 70-€ *****4.0 ○ 11 lectures ③ 33 mins | | Beginner Level III English [Auto-generated] (66 ratings)







6. Any other additional use cases (semantic pipeline, natural language processing, notifications, collecting information, stored queries, export templates, etc.)
7. Who will be the main contact person from project team to coordinate appointments etc.
For project and agenda related questions you can contact jack.mccormack@stud.h-da.de
Technical questions please forward to lucas.hammermueller@stud.h-da.de and fabian.brechtel@stud.h-da.de
8 Any existing project plan that you can share with us?
As mentioned before, we focused on API access. Our plan so far was to request the data from different sources and process it in a search-engine like "elasticsearch".
https://drive.google.com/file/d/1swrK8pYcSSOxapi5ex4l5eNVo7EJKzb-/view?usp=sharing

- 9. Which entities and structures should be extracted from structured and unstructured contents (e.g. for dynamic filters for the end-user)
 - We want to pull as much information as we can (e.g. course title,content,price,provider) and filter it
 - We focused so far on the more popular ones like https://udemy.com and https://udemy.com
- · Use Cases: Description of concrete use case and hands-on examples
 - Searching for courses from different moocs/providers
 - Presenting unified results
 - Filter the search results

.

Additional ressources:

Lufthansa User Research:

• https://docs.google.com/presentation/d/1laaD5YHNH_VnM9YopasMQOvf5XCja Clz18IU-3m_DKM/edit#slide=id.gc6fa3c898_0_0

- Motivation & Introduction / Search
- Architecture
- Web Connector
- Semantics
- Search Apps
- Operations
- Data Integration Connector