



Harvard Division of
Continuing Education



Where Did She Say It? Indexing Transcripts to Facilitate Searching


Jay Luker jay_luker@harvard.edu
Rute Santos rute_santos@harvard.edu

Why Indexing Transcripts?

- All lectures have automated transcripts generated
- Instructors can search their recordings from previous terms
- Students can review material, study for exams



Searching


Account
Admin
Dashboard
Courses
Calendar
Inbox
History
Help

BIOS E-60 (23186) > Immunology

2020-2021 Spring


- Home
- Syllabus
- Announcements
- Zoom
- Course Videos
- Assignments
- Discussions
- Quizzes
- Manage Course
- Course Evaluations
- Rubrics
- New Analytics
- Search
- App Store
- Course Admin Tools
- Student Resources
- Files
- Modules
- Pages
- Grades
- People
- Collaborations
- Outcomes
- Secure Exam Proctor
- Settings

Immunology

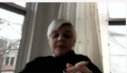
Report A Problem
More Information

Enter words or "phrase in quotes" to search...

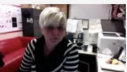
Sorted: **Newest**




Lecture 6
Mon, 08 Mar, 5:05PM
Development and survival of B lymphocytes. The adaptive immune response. T cell mediated immunity, COVID disease and host responses.




Lecture 5
Mon, 01 Mar, 4:56PM
Antigen presentation by MHCI and MHCII. Generation of T cell receptor ligands. Signaling through immune receptors. Development of T lymphocytes.




Lecture 4
Mon, 22 Feb, 5:11PM
Innate immunity (cont), Adaptive Immunity



Lecture 3
Mon, 08 Feb, 4:58PM
Toll-like receptors-structure, function, cellular expression. Inflammasomes.



Lecture 2
Mon, 01 Feb, 4:54PM
Innate Immunity. Complement system. Interferon responses



Lecture 1
Mon, 25 Jan, 4:52PM
Introduction to biology of the immune system.



Searching

Immunology

BIOS E-60 Publication Listing



Lecture 1

Mon, 25 Jan, 4:52PM

Introduction to biology of the immune system.

- with the **vaccine**, which was
- about the COVID **vaccine**
- an appropriate **vaccine**.



Lecture 4

Mon, 22 Feb, 5:11PM

Innate immunity (cont), Adaptive Immunity

- like a BCG **vaccine**, which



Lecture 2

Mon, 01 Feb, 4:54PM

Innate Immunity. Complement system. Interferon responses

- the target of current **vaccine**

BIOS E-60 (23186) > Immunology

2020-2021 Spring

- Home
- Syllabus
- Announcements
- Zoom
- Course Videos
 - Assignments
 - Discussions
 - Quizzes
 - Manage Course
 - Course Evaluations
 - Rubrics
 - New Analytics
 - Search
 - App Store
 - Course Admin Tools
- Student Resources
 - Files
 - Modules
 - Pages
 - Grades
 - People
 - Collaborations
 - Outcomes
 - Secure Exam Proctor
 - Settings



Searching

Immunology

BIOS E-60 Publication Listing

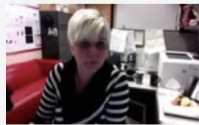


Lecture 1

Mon, 25 Jan, 4:52PM

Introduction to biology of the immune system

- with the **vaccine**, which was
- about the COVID **vaccine**
- an appropriate **vaccine**.



Lecture 4

Mon, 22 Feb, 5:11PM

Innate immunity (cont), Adaptive Immunity

- like a BCG **vaccine**, which

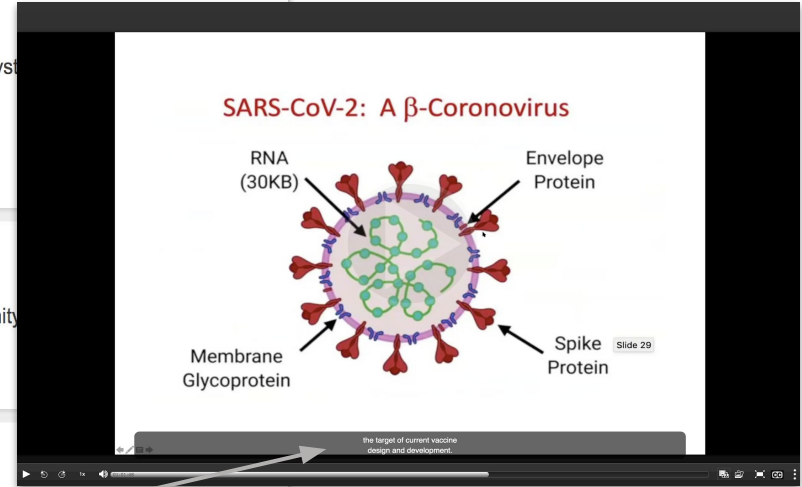


Lecture 2

Mon, 01 Feb, 4:54PM

Innate Immunity. Complement system. Interferon responses

- the target of current **vaccine**



Indexing

1. Workflow: ingest, trim, encode, publish, start automated transcripts
2. When transcripts ready, start workflow:
 - Attach transcripts to media package
 - Call AWS lambda to add to AWS Elasticsearch
3. Transcripts are searchable from video listing



Components

- Search Content service
- AWS Elasticsearch
- AWS Lambda

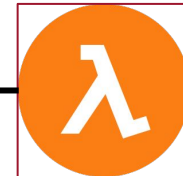


AWS Lambda Function Input



```
"message": {  
  "indexName": "oc-prod-transcripts",  
  "captionsUrl": "https://\[s3-host+path\]/captions\_.xml",  
  "mpid": "41e35f2a-e6ee-4d87-aed0-a714f7ca3e10",  
  "seriesId": "20190229997"  
}
```

```
<div xml:lang="en">  
  <p begin="0:00:01.150" end="0:00:07.750">right hello good e  
  be are this is a guest lecture </p>  
  <p begin="0:00:07.750" end="0:00:13.820">for C. S. C. I. E  
  recording this this video will </p>  
  <p begin="0:00:13.820" end="0:00:18.160">probably be re sho  
  <p begin="0:00:22.750" end="0:00:27.780">maybe that they wo  
  recording it today this is our special </p>  
  <p begin="0:00:27.780" end="0:00:34.520">guest lecture on s  
  of in check are has asked me </p>  
  ...
```



Indexing Output

- Document "_id" gets "{mpid}-{series_id}"

```
"_id": "4c36d77f-134e-459e-b8e5-5a0969f71fdf-20210224033",
```

- "mpid" and "series_id" also indexed separately as document fields

```
"_source": {  
  "mpid": "4c36d77f-134e-459e-b8e5-5a0969f71fdf",  
  "series_id": "20210224033",  
  "text": "right hello good evening welcome or welcome back is that as the case  
S. C. I. E. 40 are for November the well was 11/14/2017 when I'm recording  
are for multiple years so maybe that they won't mean anything to you but  
is our special guest lecture on spam and email filtering are something t
```

- "text" field gets a concatenation of all captions

- Each caption indexed individually as a nested document array in "captions" field

```
"captions": [  
  {  
    "text": "right hello good evening welcome or welcome back is  
    "begin": 1  
  },  
  {  
    "text": "for C. S. C. I. E. 40 are for November the well was  
    "begin": 7  
  },  
]
```



Supporting Phrase Search

- How to preserve the begin time data of the caption terms ...?
- ... but still support phrase search across captions?

```
{
  "text": "mention definitions right are this is a standard",
  "begin": 357
},
{
  "text": "definition of spam it's still pretty good",
  "begin": 362
},
```



"Good Enough" Querying

```
"bool": {  
  "must": [  
    ,  
    ],  
  }  
}
```

"standard definition"


standard definition

...before we go there I do want to talk I should just mention definitions right are this is a **standard definition** of spam it's still pretty good spam is balk messages that are unsolicited so if you send in unsolicited message you're not streaming so if I send a message to you just saying hi I sent a message to my doctor saying...

```
{  
  "text": "mention definitions right are this is a standard",  
  "begin": 357  
},  
{  
  "text": "definition of spam it's still pretty good",  
  "begin": 362  
},
```



"Good Enough" Search Results

 HARVARD UNIVERSITY

Immunology

BIOS E-60 Publication Listing

"immune response"



Lecture 7

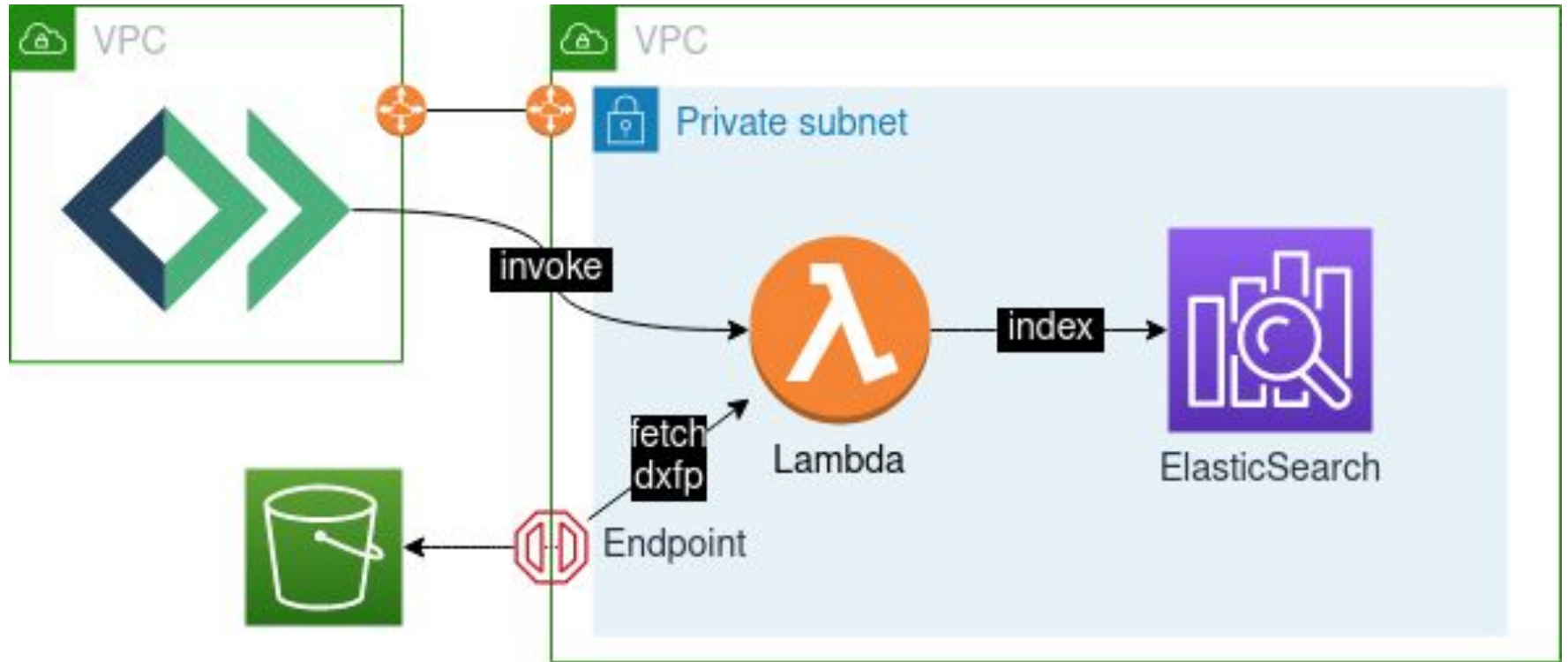
Mon, 29 Mar, 4:59PM

Humoral Immunity

- of the *immune response*.
- of the *immune response*.
- So *immunity* in which
- in the non-*immune* animal
- in a non-*immune* mouse.
- been *immunized* with antigen?
- have a decrease in their *immune*
- *responses* in animals which
- of the *response*, the
- facilitate antibody *responses*.
- is the *response* and the
- in response to a



AWS Stack



Thank You!

jay_luker@harvard.edu

rute_santos@harvard.edu

