execute ./download

Parse command line arguments

category.yml

load yaml file if necessary

Pipeline 1
download
minimum viable implementation

For each category:
1. query category

   wiki_module.query_category

2. write to db

   database_module.insert_category

3. add page to list of page ids

page_ids

For each page:
1. query page

   wiki_module.query_page
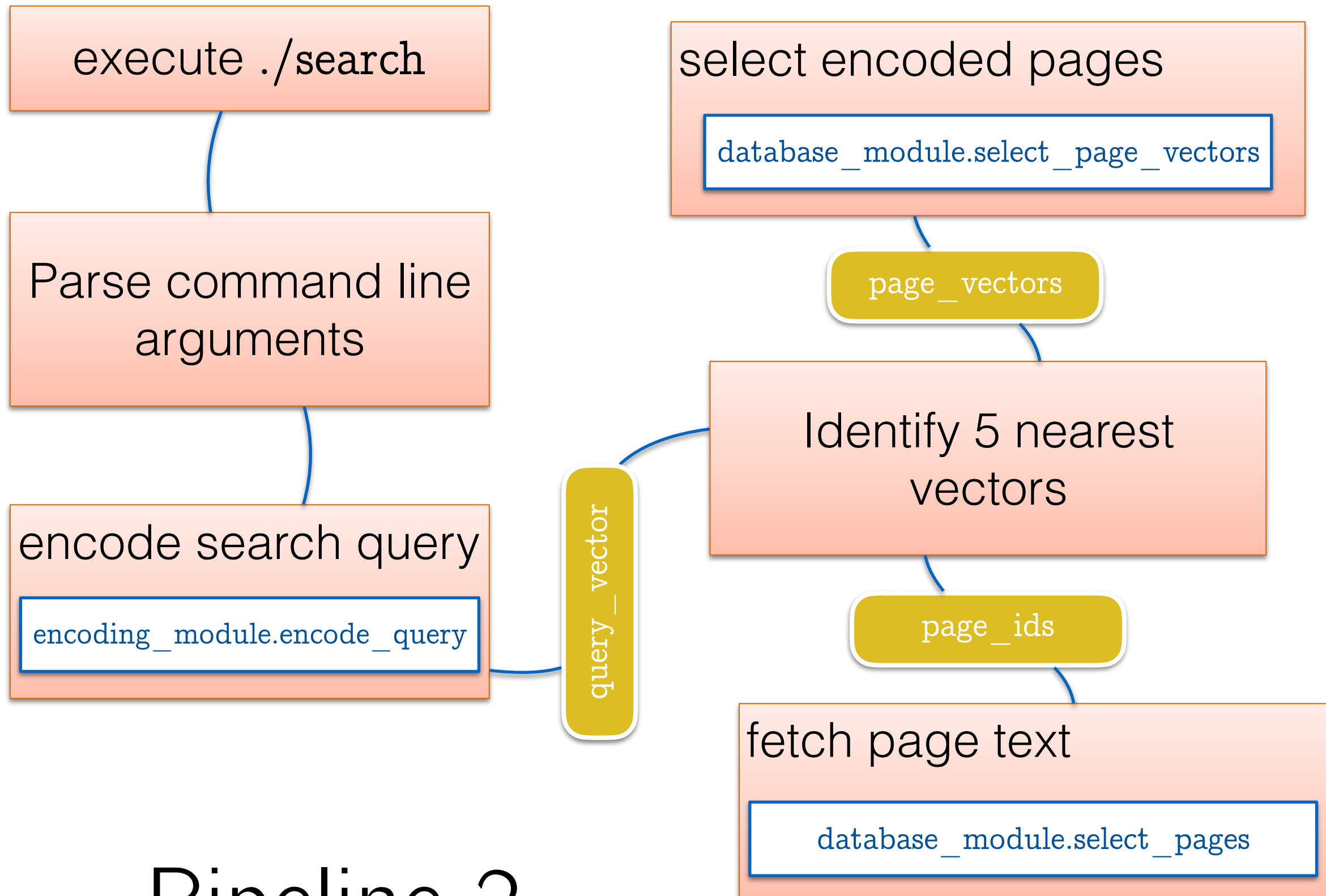
2. encode page

   encoding_module.encode_page
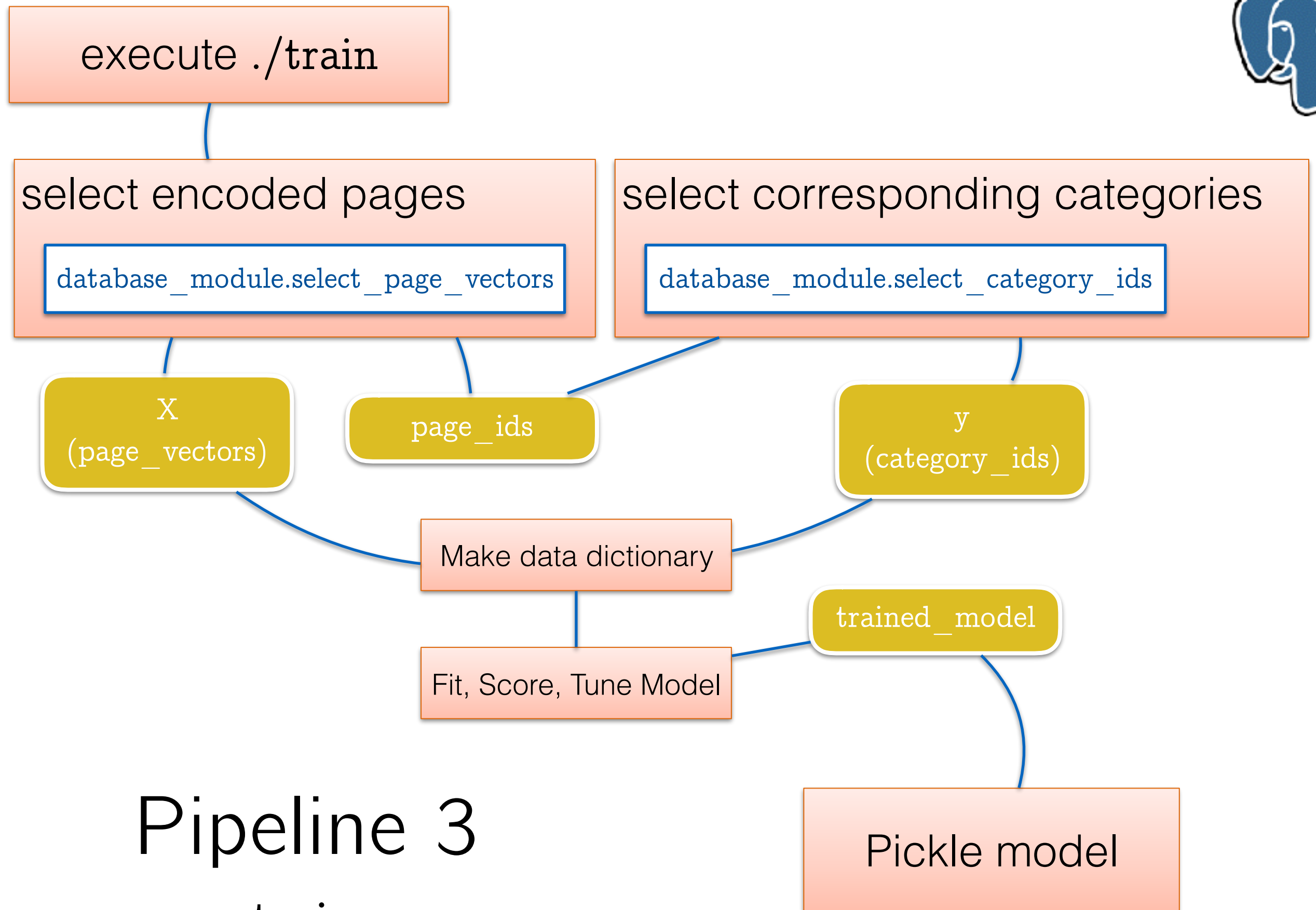
3. write to db

   database_module.insert_page

execute ./search

Parse command line arguments

encode search query

encoding_module.encode_query

query_vector

select encoded pages

database_module.select_page_vectors

page_vectors

Identify 5 nearest vectors

page_ids

fetch page text

database_module.select_pages

Pipeline 2

search

minimum viable implementation

execute ./train

select encoded pages

database_module.select_page_vectors

select corresponding categories

database_module.select_category_ids

X
(page_vectors)

page_ids

y
(category_ids)

Make data dictionary

trained_model

Fit, Score, Tune Model

Pickle model

Pipeline 3

train

minimum viable implementation

execute ./predict

query page

`wiki_module.query_page`

encode page

`encoding_module.encode_page`

load pickled model

`trained_model`

`page_vector`

predict category
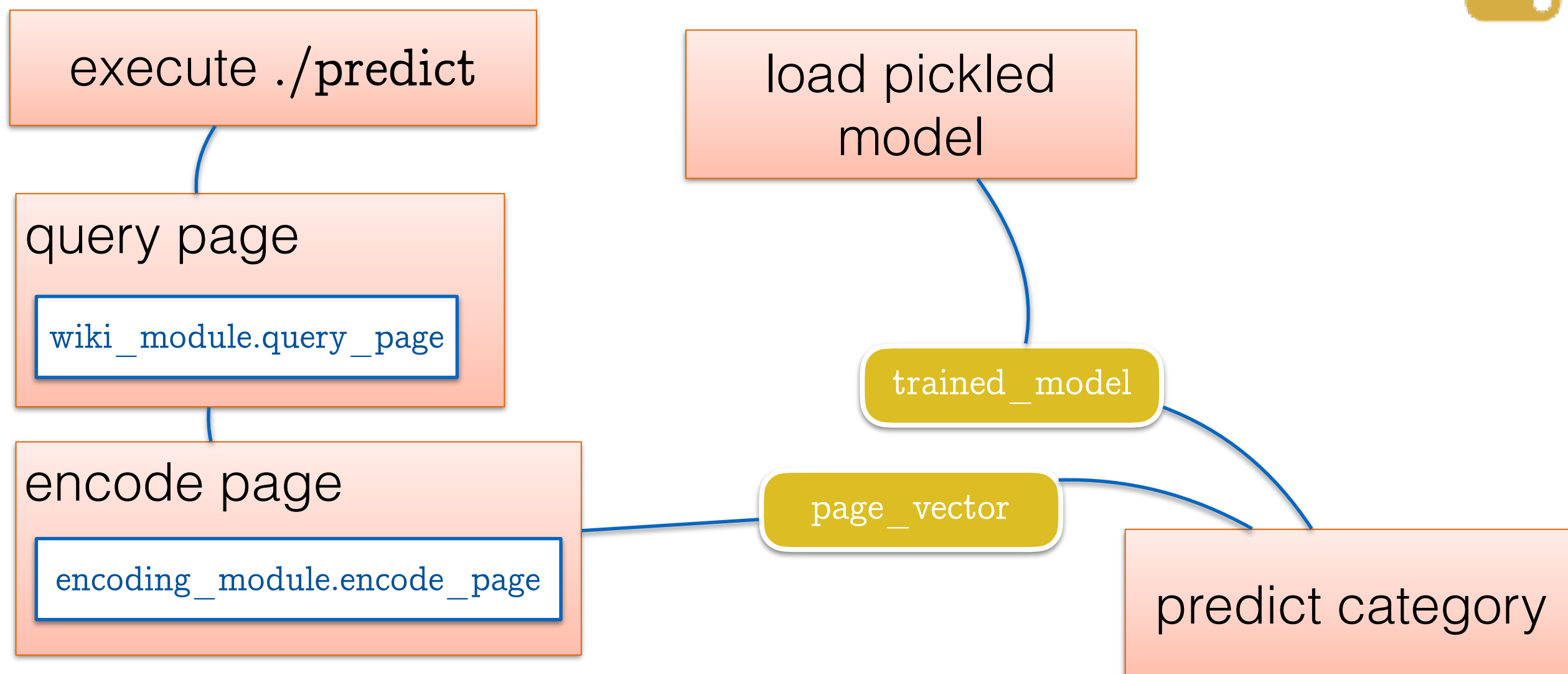
Pipeline 4

`predict`

minimum viable implementation