

# Διαδίκτυο κι Εφαρμογές

Φοίβος – Ευστράτιος Καλεμκερής (03116010)

[phoevoskal@gmail.com](mailto:phoevoskal@gmail.com)

<https://github.com/phoevos/shaka>

# ΘΕΜΑ, ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ

Η εφαρμογή βασίστηκε σε προσαρμογή της ιδέας COVID-02:

Εύρεση Άρθρων που αναφέρουν ένα συγκεκριμένο Φάρμακο είτε στην συνοπτική τους περιγραφή είτε μέσα στο πλήρες κείμενο.

Ως πηγή δεδομένων χρησιμοποιήθηκε το dataset του CORD-19 (<https://www.semanticscholar.org/cord19>), το οποίο περιλαμβάνει περίπου 200.000 άρθρα σχετικά με τον COVID-19.

Από το dataset αντλήσαμε για κάθε άρθρο τον τίτλο, τους συγγραφείς, την ημερομηνία δημοσίευσης, το abstract και το πλήρες κείμενο όπου αυτά ήταν διαθέσιμα.

Για την αποθήκευση των δεδομένων χρησιμοποιήθηκε MongoDB.

Η επικοινωνία με τη βάση πραγματοποιείται με ένα REST API υλοποιημένο σε NodeJS, με χρήση του ExpressJS για τη διαχείριση του routing.

Υλοποιήθηκε, τέλος, front-end με χρήση του ReactJS.

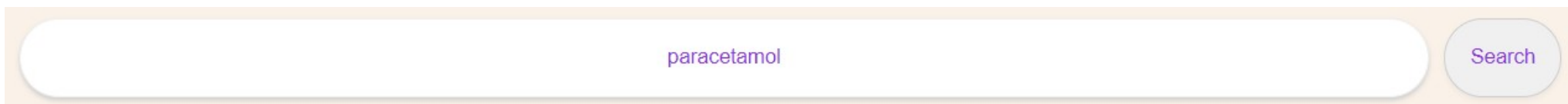
# ΛΕΙΤΟΥΡΓΙΚΟΤΗΤΕΣ

Η εφαρμογή λαμβάνει ως είσοδο το όνομα ενός φαρμάκου και:

1. Αναζητά στον τίτλο, το abstract και το πλήρες κείμενο κάθε άρθρου το συγκεκριμένο φάρμακο.
2. Επιστρέφει τα άρθρα στα οποία αναφέρεται το συγκεκριμένο φάρμακο με σειρά συνάφειας (by relevance).
3. Για κάθε άρθρο επιστρέφεται ο τίτλος, οι συγγραφείς, η ημερομηνία δημοσίευσης και το abstract, ενώ υπάρχει η δυνατότητα εμφάνισης και του πλήρους κειμένου ενός άρθρου, εφόσον αυτό είναι διαθέσιμο.
4. Εμφανίζει ιστόγραμμα με τον αριθμό των άρθρων που αναφέρουν το συγκεκριμένο φάρμακο ανά έτος, αλλά και ανά μήνα.

# ΑΝΑΖΗΤΗΣΗ ΦΑΡΜΑΚΟΥ ΣΤΑ ΑΡΘΡΑ

Η αναζήτηση φαρμάκων πραγματοποιείται από το client-side μέρος της εφαρμογής με την εισαγωγή του ονόματος ενός φαρμάκου σε ένα search bar:

A horizontal search bar with a light gray background and rounded corners. Inside the bar, the word "paracetamol" is written in a purple font. To the right of the bar is a circular button with a light gray background and a purple border, containing the word "Search" in a purple font.

Με κλικ στο Search αποστέλλεται GET HTTP Request στο API, με το όνομα του φαρμάκου να μεταφέρεται μέσω του URL.

Για τη διαχείριση της αναζήτησης σε κείμενο στη βάση έχουμε υλοποιήσει έναν text index πάνω στα πεδία text (πλήρες κείμενο), abstract και title της συλλογής Article. Επιλέχθηκε η χρήση μίας μόνο συλλογής, αντί για δύο που είχαμε σχεδιάσει αρχικά, προκειμένου να καταστεί δυνατή η ταυτόχρονη αναζήτηση και στα τρία προαναφερθέντα πεδία, αλλά και η κατάταξη των αποτελεσμάτων βάσει “σχετικότητας”. Για το τελευταίο, χρησιμοποιήσαμε το μέγεθος textScore που υπολογίζει αυτόματα η Mongo, αφού σ’ αυτό συνυπολογίζονται και ταιριάσματα στα οποία εντοπίζεται κοινή ρίζα με τη λέξη που αναζητούμε (παράγωγα, σύνθετα ή απλά διαφορετική γλώσσα, εφόσον το dataset είναι διεθνές), ενώ εφαρμόζονται και συντελεστές βαρύτητας ανάλογα με το πεδίο στο οποίο βρέθηκε η λέξη, γεγονός που το καθιστά μακράν προτιμότερο απ’ τον απλό υπολογισμό πλήρων ταιριασμάτων.

# ΑΝΑΖΗΤΗΣΗ ΦΑΡΜΑΚΟΥ ΣΤΑ ΑΡΘΡΑ

Η παραπάνω αναζήτηση επιστρέφει αποτελέσματα της μορφής:

Τίτλος **Faut-il contrôler la fièvre dans les infections sévères ?**

Abstract

Συγγραφείς Seguin, P.; Launey, Y.; Nessler, N.; Malledant, Y.

Temperature control during severe sepsis is currently used in intensive care and involves 66% and 70% of severe sepsis and septic shock, respectively. Nevertheless, the conclusive evidence of the benefit of such a strategy is still lacking. We might wonder, with regards to experimental works and recent noninterventional studies, about the risk of a control strategy on an ongoing infectious process, the patient's outcome, and the safety of the means implemented to obtain temperature control. On the other hand, it is also demonstrated that fever increases oxygen consumption, which may lead in some clinical situations to tissular ischemia and that fever may be associated with a deleterious focal inflammatory process. Methods to control the temperature include external and/or internal cooling and/or antipyretic medications such as paracetamol and nonsteroidal antiinflammatory drugs. In septic patients, external cooling and paracetamol are the main means used to control temperature. Despite the uncertainties about the benefit to control or not the temperature, it could be stated that extreme temperature (hypo- or hyperthermia) should be avoided and that the benefit/risk of temperature control must be individually weighted.

Υπαρξη πλήρους [CLICK TO READ THE ARTICLE.](#)  
άρθρου

Ημερομηνία  
2016-03-11

# ΑΝΑΖΗΤΗΣΗ ΦΑΡΜΑΚΟΥ ΣΤΑ ΑΡΘΡΑ

Στη διεπαφή χρήστη δίνεται επιπλέον η δυνατότητα πλοήγησης μεταξύ των σελίδων των αποτελεσμάτων, αλλά και φιλτραρίσματος μόνο των άρθρων που διαθέτουν πλήρες κείμενο.



dexamethasone

Search

[Click to view chart](#)

[Only show articles with available text](#)

< 1 >

## Reducing dexamethasone antiemetic prophylaxis during the COVID-19 pandemic: recommendations from Ontario, Canada

Grant, Robert C.; Rotstein, Coleman; Liu, Geoffrey; Forbes, Leta; Vu, Kathy; Lee, Roy; Ng, Pamela; Krzyzanowska, Monika; Warr, David; Knox, Jennifer

**PURPOSE:** People with cancer face an elevated risk of infection and severe sequelae from COVID-19. Dexamethasone is commonly used for antiemetic prophylaxis with systemic therapy for cancer. However, dexamethasone is associated with increased risk of viral and respiratory infections, and causes lymphopenia, which is associated with worse outcomes during COVID-19 infections. Our purpose was to minimize dexamethasone exposure during antiemetic prophylaxis for systemic therapy for solid tumors during the COVID-19 pandemic, while maintaining control of nausea and emesis. **METHODS:** We convened an expert panel to systematically review the literature and formulate consensus recommendations. **RESULTS:** No studies considered the impact of dexamethasone-based antiemetic regimens on the risk and severity of COVID-19 infection. Expert consensus recommended modifications to the 2019 Cancer Care Ontario Antiemetic Recommendations. **CONCLUSION:** Clinicians should prescribe the minimally effective dose of dexamethasone for antiemetic prophylaxis. Single-day dexamethasone dosing is recommended over multi-day dosing for regimens with high emetogenic risk excluding high-dose cisplatin, preferably in combination with palonosetron, netupitant, and olanzapine. For regimens with low emetogenic risk, 5-HT<sub>3</sub> antagonists are recommended over dexamethasone.

[CLICK TO READ THE ARTICLE.](#)

2020-06-30

# ΑΝΑΖΗΤΗΣΗ ΦΑΡΜΑΚΟΥ ΣΤΑ ΑΡΘΡΑ

Δίπλα φαίνεται η συνάρτηση που εκτελεί την αναζήτηση που περιγράψαμε προηγουμένως

*(back-end/controllers/textController.js)*

**Στη συνάρτηση έχει ληφθεί πρόνοια για pagination (skip-limit) με 10 αποτελέσματα σε κάθε σελίδα.**

**Η αναζήτηση εκτελείται πάνω στα πεδία που έχουμε συμπεριλάβει στον text index, δηλαδή στα text, abstract, title.**

```
async function get texts (req, res) {
  const drug = req.params.drug
  let result = await Article.find(
    {
      $text: { $search: drug }
    },
    {
      score: { $meta: "textScore" }, text: 0
    })
    .sort( {score: { $meta: "textScore" } } )
    .skip(req.params.page * 10 - 10)
    .limit(10)
  res.send(result)
}
```

## ΕΜΦΑΝΙΣΗ ΠΛΗΡΟΥΣ ΚΕΙΜΕΝΟΥ ΑΡΘΡΟΥ

Αν στη σελίδα με τα αποτελέσματα αναζήτησης επιλέξουμε ένα άρθρο που διαθέτει ένδειξη “Click to read full article” μεταφερόμαστε σε νέα σελίδα όπου μπορούμε να διαβάσουμε το πλήρες άρθρο. Αυτό έχει συντεθεί κατά τη διάρκεια του import των δεδομένων από το σχετικό .json αρχείο. Για την ανάσυρσή του εκτελούμε νέο GET HTTP Request, αυτή τη φορά με είσοδο το `_id` του άρθρου (το έχουμε αποθηκεύσει στο state του front-end), ενώ η σχετική συνάρτηση στο API είναι η ακόλουθη:

```
async function getTextBody (req, res) {  
  const objID = req.params.text  
  let result = await Article.findOne({_id: objID}, {abstract: 0})  
  res.send(result)  
}
```



# ΕΜΦΑΝΙΣΗ ΠΛΗΡΟΥΣ ΚΕΙΜΕΝΟΥ ΑΡΘΡΟΥ

## Η σελίδα με το πλήρες κείμενο:



[Return to search page](#)

### Reducing dexamethasone antiemetic prophylaxis during the COVID-19 pandemic: recommendations from Ontario, Canada

Grant, Robert C.; Rotstein, Coleman; Liu, Geoffrey; Forbes, Leta; Vu, Kathy; Lee, Roy; Ng, Pamela; Krzyzanowska, Monika; Warr, David; Knox, Jennifer

The COVID-19 pandemic is growing exponentially, with over two million infections and 130,000 deaths worldwide as of April 15, 2020 [1]. Early evidence suggests that patients with cancer face an elevated risk for COVID-19 infection and a higher risk of adverse events after diagnosis [2–5], potentially because of nosocomial spread and suppressed immunity. Several guidelines for pandemic era cancer care have been recently released [6–13]. These guidelines make prudent recommendations on optimizing the delivery of systemic therapy during the COVID-19 pandemic, including to: Favor oral agents over intravenous agents when efficacy and toxicity profiles are similar [6, 7]. Deliver oral agents and supportive medications directly to the homes of patients, rather than have them picked up in person at the pharmacy [6]. Favor regimens with less frequent over more frequent IV dosing when efficacy and toxicity profiles are similar [6]. Favor intermittent treatment strategies and breaks over continuous strategies when efficacy and toxicity profiles are similar and patients are stable [6, 7]. Defer the use of medications whose primary role is to reduce the risk of long-term complications such as bone disease, where appropriate [6]. Favor neoadjuvant over adjuvant treatment strategies when efficacy and toxicity profiles are similar, to facilitate the delay of surgical dates and increase the availabilities of hospital beds and ventilators [7]. Use GCSF for primary prophylaxis in high-risk regimens [7]. We convened an expert panel to formulate recommendations on modifications to antiemetic prophylaxis during the COVID-19 pandemic to protect cancer patients. We recognize the limitations of such a document, given the rapidly evolving environment and paucity of data. Prescribe the minimal effective dose of glucocorticoids (Table 1). If nausea or vomiting occurs despite the recommended regimens, increase or add non-glucocorticoid agents such as an NK1 agent or olanzapine, before increasing the glucocorticoid dose. If no nausea or vomiting occurs with prior cycles, consider further reductions in dexamethasone. No studies identified through the systematic review assessed the impact of antiemetic dosing and the risk or severity of COVID-19 in cancer patients. Our recommendations focus on minimizing glucocorticoid use for oncology patients because data support a dose-dependent association between glucocorticoids and viral and respiratory infections [15–17]. Glucocorticoids cause immunosuppression through multiple complex mechanisms, in particular by altering gene transcription of pro-inflammatory genes like interleukins and nuclear factor-kappa-B. Glucocorticoids also deplete T and B cells essential for the immune response to viruses [18]. In COVID-19, lymphopenia is common and associates with more severe disease [19], suggesting the importance of lymphocytes in the immunological response. Currently, guidelines recommend against using glucocorticoids to treat respiratory failure associated with COVID-19 in the absence of acute respiratory distress syndrome or patient-specific indications like concomitant chronic obstructive pulmonary disease exacerbation [20]. This recommendation is based on a signal of harm when using glucocorticoids to treat other viral infections.

# ΕΜΦΑΝΙΣΗ ΙΣΤΟΓΡΑΜΜΑΤΟΣ

Στη σελίδα αναζήτησης εμφανίζεται, επιπλέον, κουμπί που μεταφέρει το χρήστη σε νέα σελίδα για την προβολή του ιστογράμματος

[Click to view chart](#)

Για τη δημιουργία αυτού μας ενδιαφέρει ο υπολογισμός του αριθμού των άρθρων ανά ημερομηνία, που πραγματοποιείται με νέο GET HTTP Request που διαχειρισόμαστε με τη συνάρτηση που φαίνεται δίπλα, στο API. Αυτή, αφού εκτελέσει αναζήτηση του φαρμάκου στα κείμενα, γκρουπάρει τα αποτελέσματα βάσει της ημερομηνίας και υπολογίζει για κάθε ημερομηνία το άθροισμά τους.

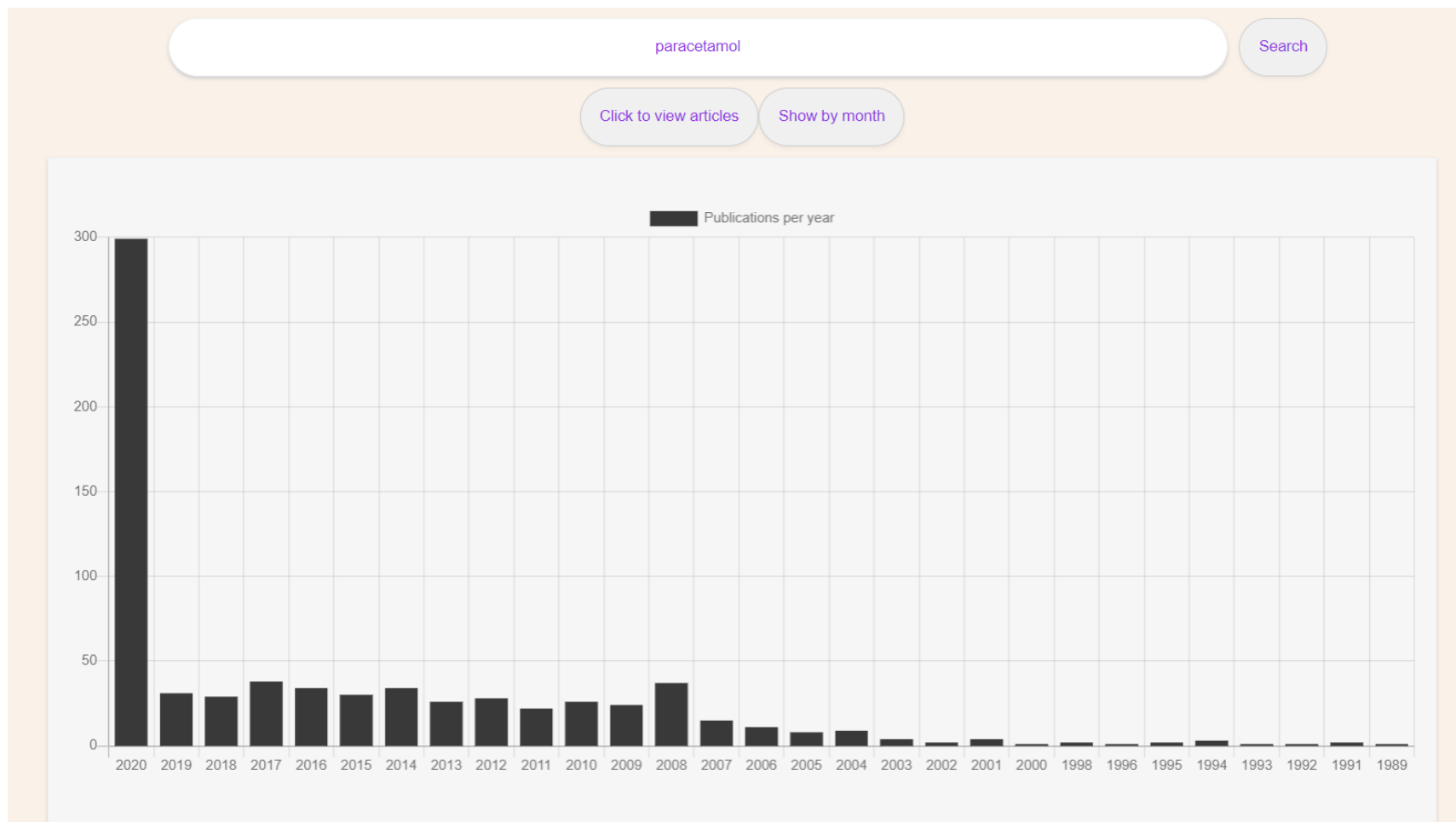
```
async function getCount (req, res) {
  const drug = req.params.drug

  let result = await Article.aggregate( [
    {
      $match : { $text: { $search: drug } }
    },
    {
      $group: {
        _id: '$publish_time',
        count: { $sum: 1 }
      }
    },
    { $sort: { _id: -1 } }
  ] )

  res.send(result)
}
```

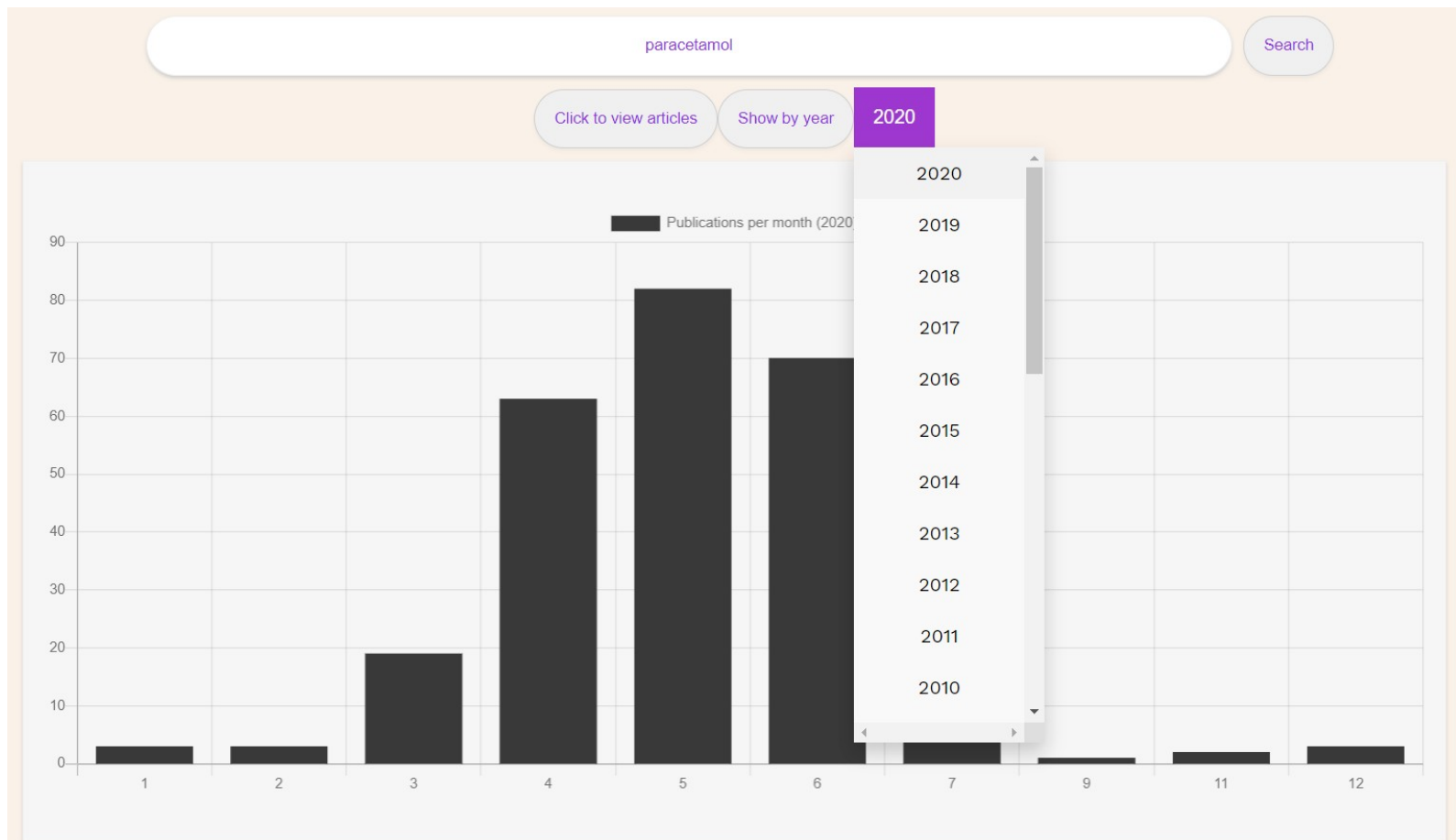
# ΕΜΦΑΝΙΣΗ ΙΣΤΟΓΡΑΜΜΑΤΟΣ “ΑΝΑ ΧΡΟΝΟ”

Συμπεριλαμβάνονται μόνο οι χρονιές κατά τις οποίες υπήρξε κάποια σχετική δημοσίευση. Με hover πάνω από κάποια στήλη μπορούμε να δούμε τον ακριβή αριθμό δημοσιεύσεων για το συγκεκριμένο χρόνο.



# ΕΜΦΑΝΙΣΗ ΙΣΤΟΓΡΑΜΜΑΤΟΣ “ΑΝΑ ΜΗΝΑ”

Στο “ανά μήνα” διάγραμμα μπορούμε να επιλέξουμε το χρόνο που μας ενδιαφέρει από μια dropdown list και να δούμε τον αριθμό δημοσιεύσεων ανά μήνα. Όπως και πριν για τις χρονιές, μήνες με μηδενικό αριθμό δημοσιεύσεων παραλείπονται.



# ΟΔΗΓΙΕΣ ΕΓΚΑΤΑΣΤΑΣΗΣ

Προτού ξεκινήσει η εγκατάσταση της εφαρμογής θα πρέπει να έχουν εγκατασταθεί τα παρακάτω:

- NodeJS (<https://nodejs.org/en/>)
- Npm (συμπεριλαμβάνεται στην εγκατάσταση του Node)
- MongoDB (<https://www.mongodb.com/try/download>)

Για να γίνει το import των δεδομένων θα πρέπει να είναι ενεργή η υπηρεσία mongod.

Προκειμένου να κατέβουν τα dependencies:

- `npm install` μέσα στο φάκελο back-end
- `npm install` μέσα στο φάκελο front-end

# IMPORT DATASET

Για το import των δεδομένων συντάξαμε το `import.js`

Αυτό κάνει parse το αρχείο `metadata.csv`, απ' το οποίο συλλέγει τον τίτλο, τους συγγραφείς, την ημερομηνία δημοσίευσης και το `abstract` για κάθε εγγραφή. Ακολουθώντας, ελέγχεται αν υπάρχει σχετικό αρχείο εντός του φακέλου `document_parsers`. Δίνουμε προτεραιότητα στα `pmc_json`, αφού σύμφωνα με το FAQ του CORD-19 είναι πιο καθαρά από τα `pdf_json`. Αφού βεβαιωθούμε ότι το αρχείο υπάρχει, δημιουργούμε `stream` από το περιεχόμενό του (αντί να το φορτώσουμε ολόκληρο στη μνήμη) και συνενώνουμε τα τμήματα του `text`. Τέλος, προσθέτουμε τα παραπάνω σε μια εγγραφή στη συλλογή `Article` της βάσης.

Από το φάκελο του back-end: `node ./import/import.js`

**ΣΗΜΑΝΤΙΚΟ:** Για μεγάλα dataset θα πρέπει να αυξήσουμε το όριο μνήμης που διαθέτουμε στο Node (ακόμα και με τη χρήση του `stream`, χρειάστηκα περίπου 5GB RAM προκειμένου να φορτώσω το dataset των 200.000 εγγραφών) ως εξής:

```
node --max-old-space-size=5120 ./import/import.js
```

# ΔΟΚΙΜΑΣΤΙΚΟ DATASET

Στο φάκελο του project, εντός του import, έχουμε τοποθετήσει ένα δοκιμαστικό dataset που αποτελείται από περίπου 1.500 εγγραφές στο αρχείο metadata.csv, καθώς και από περίπου 60 αρχεία pdf\_json/rmc\_json, προκειμένου να διευκολύνουμε τη δοκιμή της εφαρμογής.

Για το συγκεκριμένο δοκιμαστικό dataset προτείνονται προς αναζήτηση τα εξής φάρμακα: paracetamol, dexamethasone, hydroxychloroquine.

*Σε περίπτωση που είναι επιθυμητή η χρήση του πλήρους dataset, θα πρέπει πριν το import να έχει τοποθετηθεί χειροκίνητα εντός του φακέλου import, διατηρώντας την υπάρχουσα δομή φακέλων (περιγράφεται και στο readme του back-end) και ρυθμίζοντας, όπως αναφέρθηκε, την παρεχόμενη μνήμη (για 200.000 εγγραφές απαιτούνται περίπου 5GB και 12 λεπτά)*

# ΕΚΚΙΝΗΣΗ ΕΦΑΡΜΟΓΗΣ

Για εκκίνηση της εφαρμογής:

- `node server` εντός του φακέλου `back-end`
- `npm start` εντός του φακέλου `front-end`