

闘病記ブログから患者の医療イベントを インクリメンタルに要約した データセット

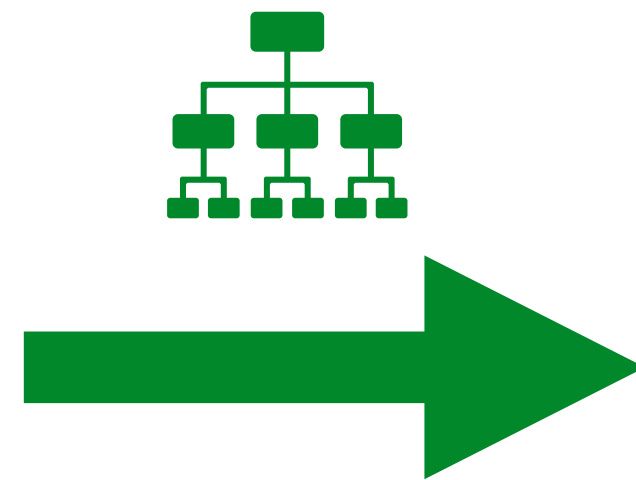
矢田 竣太郎, 荒牧 英治
奈良先端科学技術大学院大学
JLR2024 – 2024-03-15

目的 | 患者の語りから闘病過程を抽出



闘病記ブログ

スクレイピング

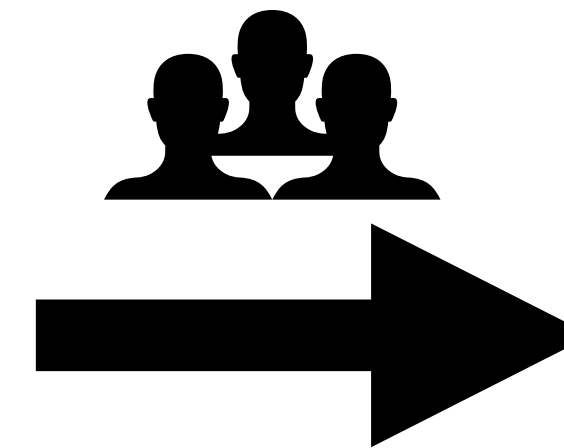


4K乳がん闘病記ユーザ
732K記事

11月3日、ついに入院生活が始まりました。
翌日、最初の放射線治療です。
(.....)
入院4日目、早くもつらくなってきました。吐き気で食欲もわかりません。
(.....)
いよいよ7回目の放射線治療です。これで1/5が終わります。
(.....)
先週末あたりから舌の痛みが気になるようになってきました。
(.....)

闘病記コーパス

人手*で要約



100ユーザをサンプル**
5743記事

20XX年

11月3日 入院開始

11月4日 放射線治療1回目
(.....)

11月7日 吐き気、
食欲わかず
(.....)

11月Y日 放射線治療7回目
(.....)

11月Z日 舌の痛みを感じ始める

医療イベント要約文

*医療従事者2名に依頼

**総記事数10~200かつ総文字数20万以下のユーザからランダムサンプル 2

応用例 | 時系列可視化

11月3日、ついに入院生活が始まりました。
翌日、最初の放射線治療です。
(……………)
入院4日目、早くもつらくなってきました。吐き気で食欲もわきません。
(……………)
いよいよ7回目の放射線治療です。これで1/5が終わります。
(……………)
先週末あたりから舌の痛みが気になるようになってきました。
(……………)

医療
イベント
要約

20XX年
11月3日 入院開始
11月4日 放射線治療1回目
(……………)
11月7日 吐き気,
食欲わかず
(……………)
11月Y日 放射線治療7回目
(……………)
11月Z日 舌の痛みを感じ始める

臨床医学
アノテーション
+
時間関係推論

例)
11/29 結節に uptakeなし, フォローを.
TIMEX3(DATE) d(+) c CC(?)



時系列 (タイムライン)
可視化



ブログ



図書

闘病記テキスト

医療イベント要約文

Sparseな患者記述をDenseな箇条書きに

既存手法はDenseな医師記述に特化

応用例 | 患者向けチャットボット



患者

もう朝だね。
結局**眠れなかった**よ
おなかが空いたんだけど、作るのがめん
どくさかったからお湯をそそぐだけの
カップ焼きそばを食べる。
でも、その前に。。。
食道が痛い。。。
なんか飲み込むだけで痛い。。。

朝ですね。眠れなかったのはつらいです
ね。カップ焼きそば、手軽でいい選択です
ね。ただ、食道が痛いのは心配ですね。



こんな時間に**気持ち悪い**。
吐き気止め効かない。
血圧が100代から70代に**一気に下がっ**
た。
布団に横になると**めまい**がする。

こんな時間に薬の効かない吐き気がある
とつらいですね。しかも血圧が急に下がっ
てめまいがするのは心配だと思います
が、先生と相談しますか？

チャットボット



20XX年

11月3日 入院開始
11月4日 放射線治療1回目
(.....)
11月30日 一時退院
(.....)
12月17日 不眠
食道に痛み・嚥下痛

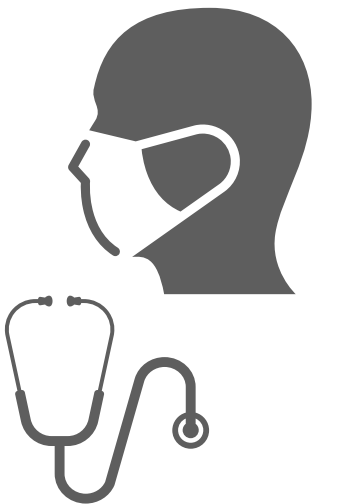
Incremental
Update



20XX年

11月3日 入院開始
11月4日 放射線治療1回目
(.....)
11月30日 一時退院
(.....)
12月17日 不眠
食道に痛み・嚥下痛
(.....)
12月23日 夜間に嘔気
処方箋効果なし
低血圧 めまい

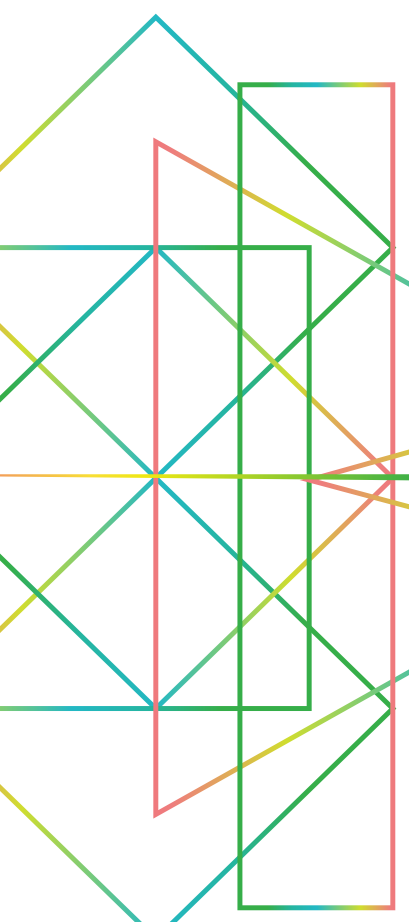
医療従事者



created_at	title	text	summary	incremental_summary
.....
2017-12-22	癌告知	検査結果の日名前を呼ばれて、診察室に入るとその気配で、やっぱり癌かあーと少しは違うかもと期待してたけど、、、先生も担当の看護師が来るので、待って下さいと。先生の隣に看護師いてるやん。えーなんで看護師待つん？担当ってどういう事？もう癌って言ってるのと同じやんで、先生からの癌告知！！先生はとても優しく、私の様子を見ながら説明してくれました。この検査結果の間最悪を考えて病院を選んだので、 転院を伝えました。すぐに転院先の予約を入れました。	がん告知約1年前 転院の予約	2012年ごろ 右胸出血 乳がん検査（エコー・マンモ） それから半年後、一年後も実施 陰性 2017年初頭 針生検 約1年前 しこり3箇所 乳がん判明 転院の予約 2月温存手術 現在ホルモン治療中
2017-12-24	転院	1月27日に新しい病院にプレパレートを持って診察。 優しい先生で、MRIや心電図の説明があって、骨シンチはしなくていいと。再発した時に、シンチをしましょうと再発？？？初期の癌だから、取ったら大丈夫と思ってた。じゃあ再発しても、また手術したら治ると思ってた。今思うと自分の無知が恥ずかしかつたで、 手術は2月15日に決定 しました。癌で有名な病院で、建物が古くお化けが出そうだから、絶対個室は無理と思ってる私です入院に向けて髪の毛をバツサリ！	2017年1月27日 転院 診察 同年2月15日に手術 予定	2012年ごろ 右胸出血 乳がん検査（エコー・マンモ） それから半年後、一年後も実施 陰性 2017年初頭 針生検 約1年前 しこり3箇所 乳がん判明 2017年1月27日 転院 診察 2月15日に温存手術 現在ホルモン治療中
.....

	min	median	max
記事数/ユーザ	10	51	198
文字数/記事	1	365	3318
最終要約字数/ユーザ	1	832	4729

インクリメンタル要約： 当該記事+過去記事の内容を要約
→ 常に最新の全要約を得たいタスク
単純な「**追記**」だけではなく過去情報の「**編集**」も生じうる



モデル構築時に気になること

- 図書1冊, ブログユーザ1名を丸ごと要約 → 長文解釈モデルが必要?
 - 入力トークン数やコンテキストウィンドウの大きいモデルで全記事を読む
- 医療応用なので運用環境はプライバシーに要配慮 → 軽量要約モデル?
 - 「各記事の要約」の要約 (Divide & Conquer) で十分解ける?
 - 「直前記事までの要約+今回記事」を要約すれば良い?
 - これまでの記事を外部記憶として保持しつつ, 今回記事を要約する試み
 - ♦ **NLP2024 A9-4**: Recurrent Memory Transformer for Incremental Summarisation of Extremely Long Text (Li, Yada, & Aramaki)