

بررسی ویژگی‌های بیماران و پزشکان در سرویس‌های پرسش و پاسخ



دانشجو: سهند عصری
استاد راهنما: دکتر مسعود رهگذر
دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

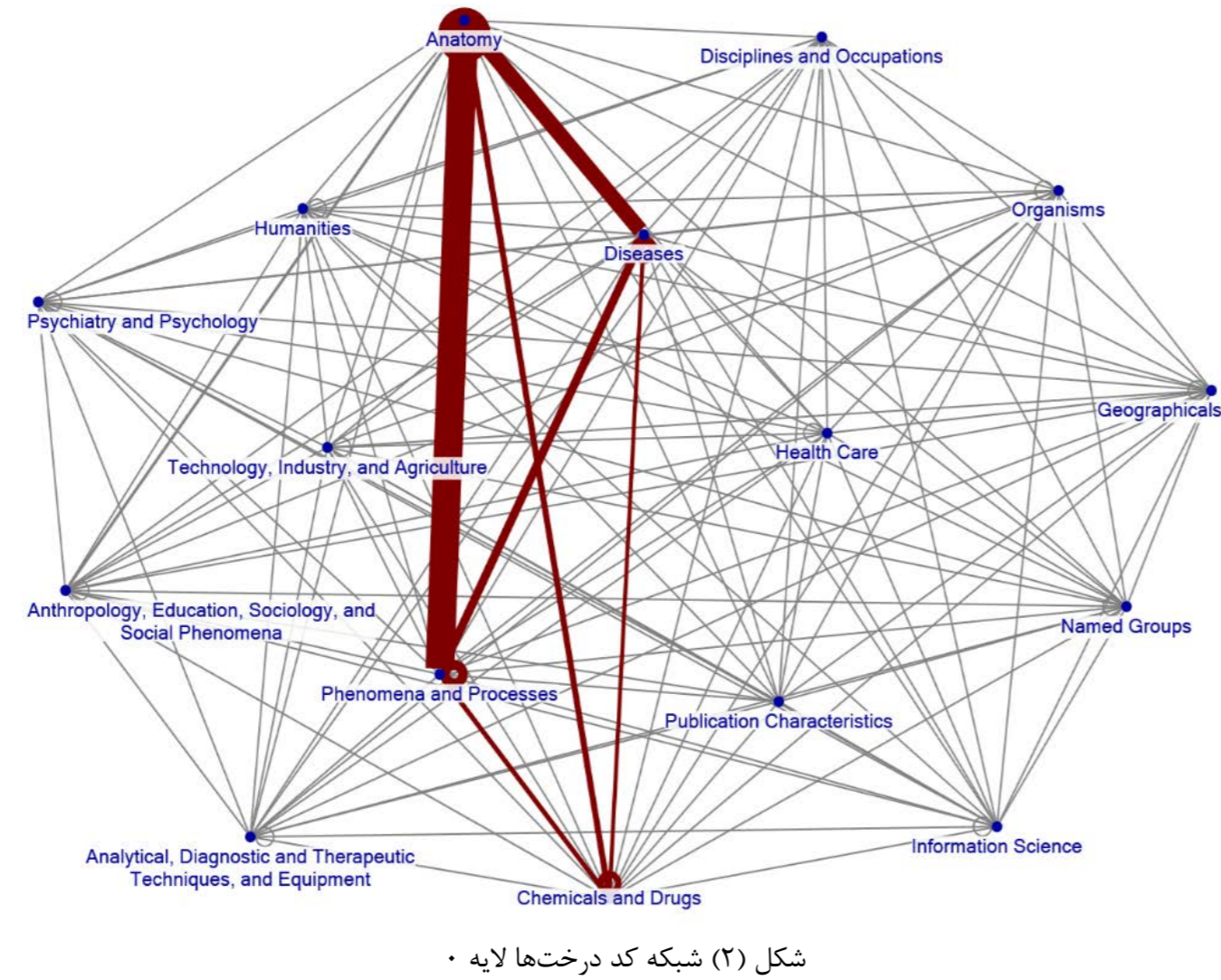


جدول (۳) پرتکرارترین نوع بیماری‌ها (C)

Treecode1	Description	CNT
C17	Skin and Connective Tissue Diseases	6,055
C14	Cardiovascular Diseases	5,498
C06	Digestive System Diseases	5,426
C20	Immune System Diseases	5,178
C10	Nervous System Diseases	3,701
C13	Female Urogenital Diseases and Pregnancy Complications	3,625

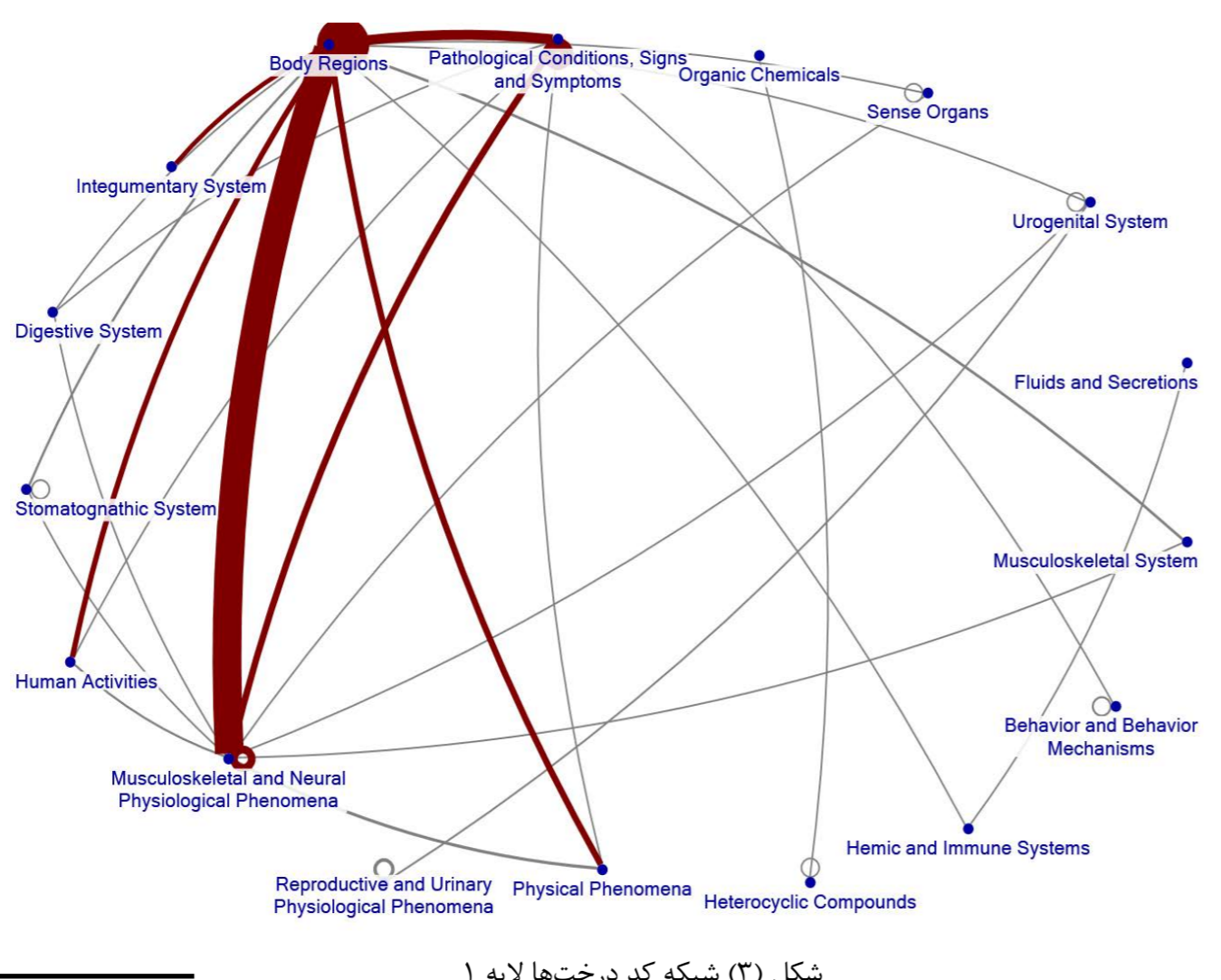
جدول (۴) کد درخت‌های زیرمجموعه آناتومی بدن (A)

Treecode1	Description	PCT
A05	Urogenital System	15.13%
A15	Hemic and Immune Systems	13.40%
A14	Stomatognathic System	13.30%
A03	Digestive System	13.24%
A09	Sense Organs	11.75%



جدول (۵) پرتکرارترین اعضای بدن (A01)

Treecode2	Concept	Count
A01.923	Back	18,637
A01.456	Head	10,717
A01.598	Neck	8,066
A01.456	Face	7,489
A01.378	Knee	4,559



جدول (۶) پرتکرارترین علائم و نشانه‌ها

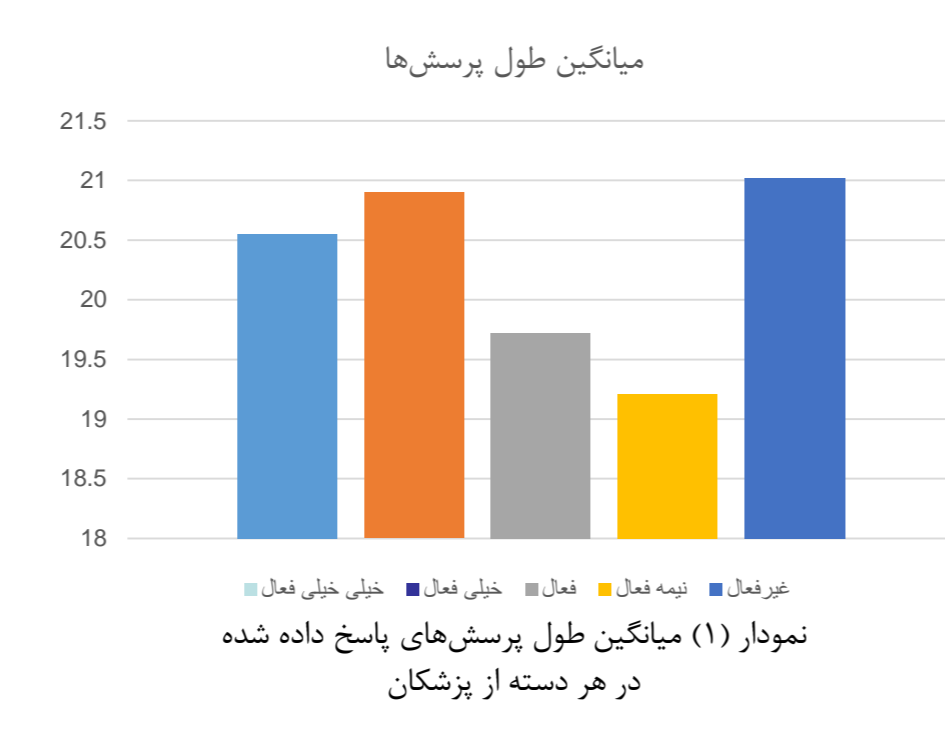
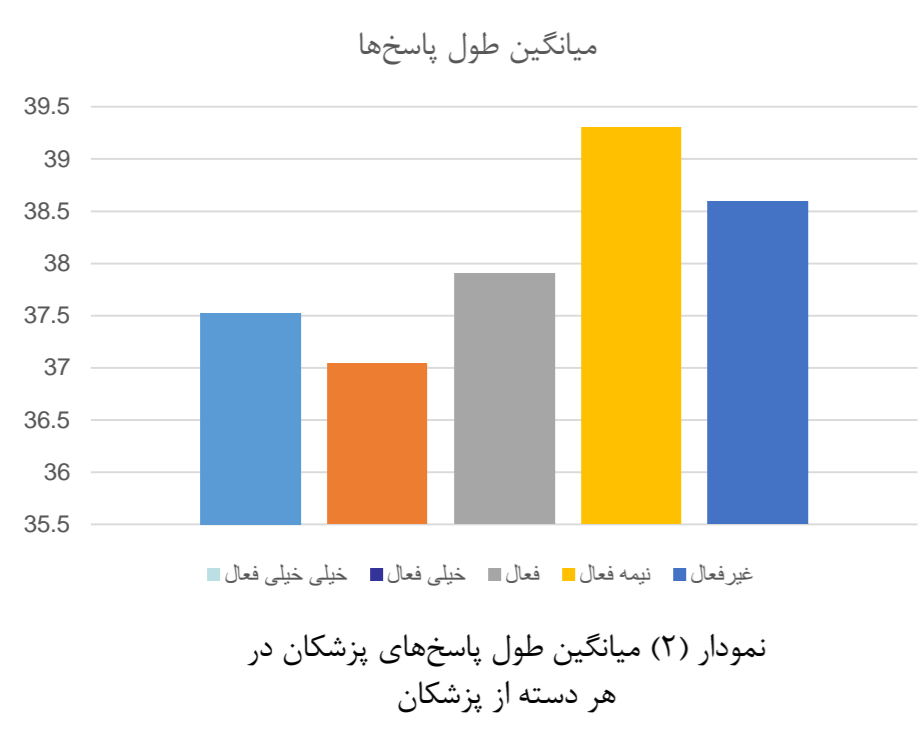
Treecode2	Description	Concept
C23.888	Signs and Symptoms	Fever
C23.888	Signs and Symptoms	Nausea
C23.888	Signs and Symptoms	Cough
C23.888	Signs and Symptoms	Fatigue
C23.888	Signs and Symptoms	Dizziness

جدول (۷) پرتکرارترین بیماری‌ها

Treecode	Description	Concept	CNT
C20.543	Hypersensitivity	Asthma	3,011
C19.874	Thyroid Diseases	Hypothyroidism	1,157
C17.800	Skin Diseases	Eczema	1,126
C23.300	Pathological Conditions, Anatomical	Hernia	1,108
C14.907	Vascular Diseases	Stroke	1,079
C13.351	Female Urogenital Diseases	Endometriosis	1,010

جدول (۸) پرتکرارترین مفاهیم رابطه A01-G11

G11	A01	CNT
Pain	Back	2,596
Pain	Neck	13,93
Pain	Head	1,261
Pain	Knee	1,147
Pain	Foot	879



کاربرد های صنعتی:
از نتایج به دست آمده از این پروژه می‌توان در ساده‌سازی و بهینه‌سازی سرویس‌های پرسش و پاسخ و همچنین روتینگ پزشک‌ها به سوال‌ها (بسته به اینکه معمولا چه نوع سوال‌هایی را پاسخ می‌دهند) استفاده کرد.

مراجع اصلی

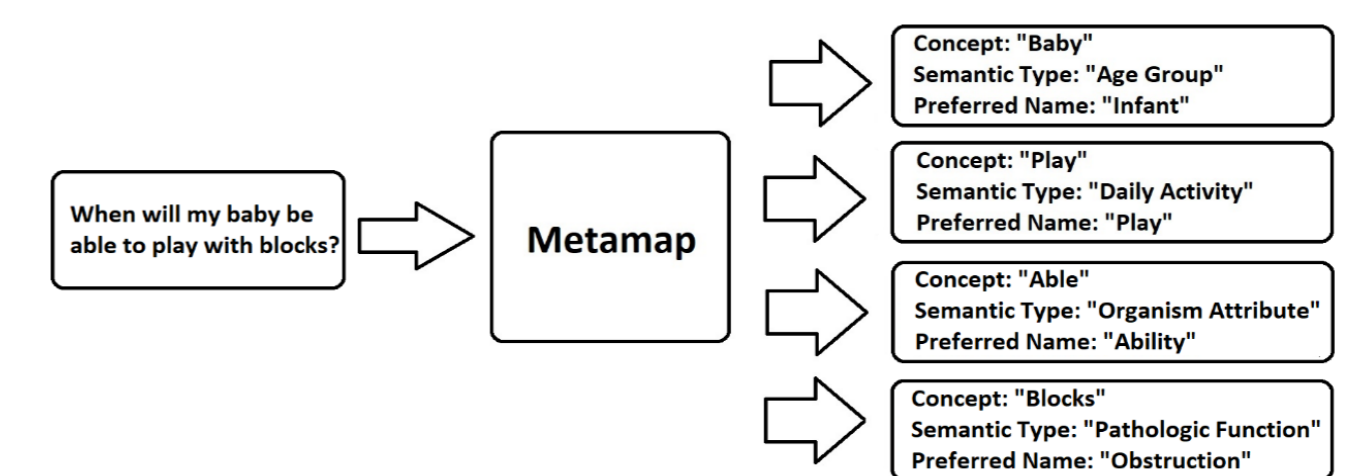
[1] Y. Zhang, "Contextualizing consumer health information searching: an analysis of questions in a social Q&A community," in Proceedings of the 1st ACM International Health Informatics Symposium, 2010, pp. 210-219: ACM.
[2] J. W. Ely et al., "A taxonomy of generic clinical questions: classification study," Bmj, vol. 321, no. 7258, pp. 429-432, 2000.
[3] M. H. P. Chiu and C. C. Wu, "Integrated ACE model for consumer health information needs: A content analysis of questions in Yahoo! Answers," Proceedings of the Association for Information Science and Technology, vol. 49, no. 1, pp. 1-10, 2012.
[4] C. Treude, O. Barzilay, and M.-A. Storey, "How do programmers ask and answer questions on the web?: Nier track," in Software Engineering (ICSE), 2011 33rd International Conference on, 2011, pp. 804-807: IEEE.
[5] National Library of Medicine, 2009. UMLS reference manual.

چکیده

در سرویس‌های پرسش و پاسخ مبتنی بر جوامع، اطلاعات ارزشمندی در قالب سوال و جواب نهفته است که می‌تواند بهبود قابل ملاحظه‌ای در ارتباطات میان پرسش‌کننده و پاسخ‌دهنده ایجاد کنند. مسئله اصلی این پروژه کشف و جمع‌آوری این اطلاعات با بررسی دقیق سوال‌ها و جواب‌ها در جنبه‌های گوناگونی از جمله ساختار نحوی و معنایی سوال و جواب‌ها می‌باشد. برای مثال بررسی تعداد کلمات پرسش‌های مطرح شده یک تخصص خاص، بررسی روابط واژگان و مفاهیم نهفته در پرسش‌ها و بررسی پرتکرارترین بیماری‌ها و علائم.

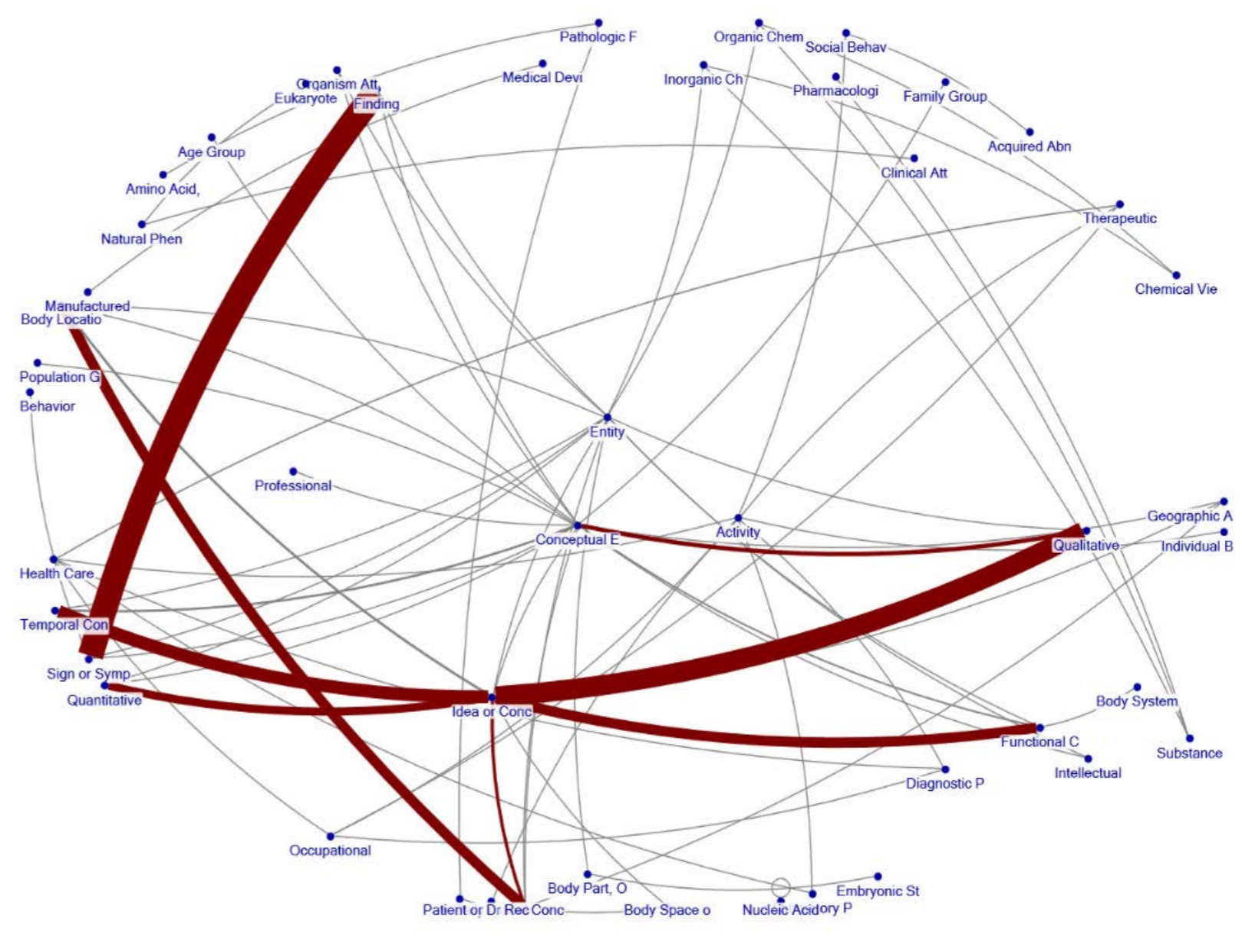
روش کار

در این پروژه، ابتدا متن پرسش‌ها و پاسخ‌های پزشکی را از وبسایت HealthTap جمع‌آوری کردیم. در ادامه، برای گردآوری مفاهیم نهفته در این متون از ویژگی‌های سیستم یکپارچه زبان پزشکی (UMLS) و یکی از پرتکرارترین ابزارهای Metamap، استفاده کردیم. متن خام پرسش‌ها و پاسخ‌ها را به عنوان ورودی به ابزار Metamap دادیم و در پاسخ عبارات با ویژگی‌هایی از جمله مفهوم، گونه معنایی و نام تجاری عبارت را دریافت کردیم.



نتایج

از جمله نتایج به دست آمده از این پروژه می‌توان به موارد زیر اشاره کرد:
بررسی پرتکرارترین گونه‌های معنایی، کد درخت‌ها، بیماری‌ها، علائم و تومورها.
بررسی پرتکرارترین گونه‌های معنایی، کد درخت‌ها، بیماری‌ها و علائم هر تخصص پزشکی
بررسی پرتکرارترین روابط میان گونه‌های معنایی و همچنین کد درخت‌های لایه ۰ و ۱
بررسی میانگین تعداد کلمات پرسش‌ها و پاسخ‌ها در هر دسته از پزشکان



جدول (۱) پرتکرارترین روابط بین گونه‌ها

Semantic Type 1	Semantic Type 2	Relation Type	PCT
Finding	Sign or Symptom	ISA	53.28%
Qualitative Concept	Idea or Concept	ISA	41.80%
Idea or Concept	Temporal Concept	ISA	27.93%
Body Location or Region	Spatial Concept	ISA	23.20%
Functional Concept	Idea or Concept	ISA	22.75%

جدول (۲) بیشترین روابط بین مفاهیم یافته‌ها و علائم

Concept 1	Concept 2	Count
Pain	Swollen	821
Cough	Fever	746
Itchy	Red	745
Cramps	Pregnant	632
HEAD ACHE	Fever	605