Question Answering in Biomedicine

Pierre Zweigenbaum

Mission de Recherche en Sciences et Technologies de l’Information Médicale, STIM/DPA/DSI, Assistance Publique - Paris Hospitals, Paris & ERM 202, INSERM

pz@biomath.jussieu.fr

http://www.biomath.jussieu.fr/~pz/

STIM ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS
Outline

Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion

General orientation: Web-based Medical QA
Health is one of the most frequent search topics on search engines

Who asks (bio)medical questions?

➢ Patients (the general public)
➢ Medical students
➢ Health care professionals
  family doctors (GPs)
➢ Biomedical researchers
➢ …
The General Public

➢ Before or after seeing a doctor

➢ For oneself or for relatives

➢ All about a given (suspected or diagnosed) disease,
  All about a given symptom,
  Indications and contraindications of treatment, etc.

Sometimes better informed on that topic than the doctor
Medical Students

- Preparing tests
- Doing assignments

Increasing amount of teaching material online at medical schools

Note: French “Internat” questions: learn answers by heart, no use for QA
Health Care Professionals

Half-life of medical knowledge is about 7 years
(2 years for biomedical knowledge)

➢ Continuous medical education
➢ Bibliographic databases (Medline, etc.)
➢ Direct Web search
Health Care Professionals: Example Questions

From (Ely et al., BMJ 2002)

- What is the dose of metformin?
- What is the proper treatment of gastro-oesophageal reflux disease (GERD)?
- What should I use for atopic dermatitis?
- How common is depression after infectious mononucleosis?
- What is the name of that rash that diabetics get on their legs?
- Is it ethical for me to take care of my own file clerk, who has back pain and wants a work excuse?
 Physicians spend less than 2 minutes on average seeking an answer to a question. (Ely et al., BMJ 1999)

Thus, most clinical questions remain unanswered. (Alper et al., J Fam Pract 2001)

Doctors are overwhelmed by the amount of information available, yet they often cannot answer their questions about specific clinical problems. (Ely et al., BMJ 2002)

A niche for question-answering technology?
Main obstacles to answering doctor’s questions about patient care (Ely et al., BMJ 2002)

➢ The **excessive time** required to find information

➢ Difficulty modifying the original question

➢ Difficulty selecting an **optimal strategy to search** for information
decide which resources will be most helpful; search in which order; which articles to read thoroughly; how thoroughly

➢ Failure of a seemingly appropriate resource to cover the topic

➢ Uncertainty about how to know when **all the relevant evidence has been found** so that the search can stop

➢ Inadequate **synthesis of multiple bits of evidence** into a clinically useful statement (e.g., conflicting evidence)
Biomedical Researchers

➢ Search structured databases (FlyBase, SwissProt, etc.),

➢ Bibliographic databases (Medline, etc.),

➢ But also the Web
  faster update, more comprehensive
Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion
**Language Specialization: Technicity**

A spectrum of levels of technicity, which rely on diverse levels of medical knowledge:

- ✗ Specialist’s medical jargon (from specialist to specialist)
- ✗ General medical talk (from/to non-specialist)
- ✗ Patient vocabulary (“consumer vocabulary”)
  - *e.g.*, Medline Plus, CISMeF patient
- ✤ Translate medical terms to patient terms (“synonym” approach)
- ✤ Use medical terms and define them for patients (“glossary” approach)
- ✤ Use different ontology for patients (Medline Plus)
Language Specialization: Terminological Variation

✗ Variant denominations for a given disease or sign

✗ Rich vocabulary: English terms, Neoclassical terms (Greek and Latin), eponyms, abbreviations...

black lung disease  
colour vision deficit  
colour vision impairment  
colour blind disease  
colour vision decline

coal miners’ lung  
miners’ lung  
melanedema  
pneumomelanosis  
col miners’ pneumoconiosis  
col miners’ pneumoconiosis  
col miners’ pneumoconiosis

coal workers’ lung  
miners’ lung  
melanedema  
pneumomelanosis  
col miners’ pneumoconiosis  
col miners’ pneumoconiosis  
col miners’ pneumoconiosis

miners’ asthma  
miners’ lung

anthracosis  
pneumomelanosis

lung melanosis  
col miners’ pneumoconiosis

coal workers’ pneumoconiosis  
col miners’ pneumoconiosis

Colliers’ lung  
Colliers’ anthracosis
Impact on search for information: example

✗ sciatica is less conducive to a literature search than low back pain

(Ely et al., BMJ 2002)
**Trust: Reliability of Sources of Medical Information**

Information used to make the most personal decisions, i.e., those concerning an individual’s health, should be of the highest quality. *(Hersh et al., JAMA 1998)*

✗ Information about date and authorship is vital to assess the quality and currency of sources

✗ Even more stringent issue for QA since direct answers can be construed as implying a higher level of endorsement
Trust: Reliability of Sources of Medical Information

✗ The bulk of information on the World Wide Web [...] is of low applicability and poor quality for answering clinical questions.
(Hersh et al., JAMA 1998, based on 50 questions)

✗ Reliability is not taken into account by QA systems
(Lin J., EACL, 12/4/2003, answer to question from the audience)
Selection bias of the Web

- Computer science
- Trivia (geography and tourism, etc.)
- Health
- Health, reliable
- Health, French
Low Availability

✗ Web redundancy may be much lower for technical, health questions

✗ Some questions have no answer on the Web

(Jacquemart & Zweigenbaum, Stud Health Technol Inform 2003)
found no answer through Google for 40% of a set of 100 student questions (oral pathology, French)

✗ Some (useful) questions have no known answer

Is smoking a risk factor for sinusitis? (Ely et al., BMJ 2002)
Need for Inference?

✗ Depends on distance between question and answering document

✗ Needs to be investigated
Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion
Quality-controlled health gateways

**CISMeF** (Catalog and Index of French-language Health Resources), Rouen University Hospital, France (*www.chu-rouen.fr*)

*(Darmoni & Thirion, Meth Inf Med 2000)*

**CliniWeb** Oregon Health Sciences University, USA

(*www.ohsu.edu/cliniweb*) *(Hersh et al., AMIA Annu Symp 1999)*

**DDRT** (Diseases, Disorders and Related Topics, *www.mic.ki.se/Diseases*), Medical Library and Medical Information Center, Karolinska Institute, Stockholm, Sweden
**Health Information Sources: Health Gateways**

**HON**  (Health on the Net Foundation, Switzerland) ([www.hon.ch](http://www.hon.ch))

(Boyer et al., *Int J Med Inf* 1997)

**MedWebPlus**  USA  ([www.medwebplus.com](http://www.medwebplus.com))

**OMNI**  (Organizing Medical Networked Information UK) ([omni.ac.uk](http://omni.ac.uk))

(Norman, *Med Inf* 1998)
Health Information Sources: CISMeF

CISMeF (Darmoni et al., Meth Inf Med 2000)

Catalog and Index of French-language Health Resources

- Yahoo-type directory; 11,000+ “resources”
- Primary intended audience: health care professionals but also general public, medical students
- Selects resources according to quality criteria www.medcertain.org
  - Addresses the issue of trust
Health Information Sources: CISMeF

- Indexes resources with structured thesaurus (MeSH: Medical Subject Headings)
  - Helps identify more precisely relevant documents
- Expanded resource type classification
  - practice guidelines; patient association; teaching material
  - Helps identify the most appropriate resources
- Manual indexing (updated weekly)
Health Information Sources: Virtual Medical University

Primary intended audience: medical students

- Online teaching material:
  - handbooks
  - exams and solutions
  - interactive pedagogical material

- e.g., UMVF consortium of French medical universities (Le Beux et al., Inf Sant 2002)
Primary intended audience: health care professionals

- NLM’s Medline (www.nlm.nih.gov)
- Increasingly large number of free, full-text, online articles (Pubmed Central, BMJ, etc.)
Primary intended audience: **health care professionals**

- most up-to-date, “evidence-based” clinical knowledge
- oncology, etc.
- mostly textual knowledge bases (*e.g.*, ANAES)
**Health Information Sources: Consumer-Oriented Sites**

Primary intended audience: general public

- Governmental sites (e.g., NLM’s MedlinePlus, Québec’s Health Information site)
- Associations of patients or families (e.g., Alzheimer’s disease)
- Commercial sites (Doctissimo, etc.)
Primary intended audience: health professionals; general public

→ Generally on CDROM; if Web, restricted access

→ Authoritative knowledge

☞ Drug knowledge bases (e.g., Vidal, Physician’s Desk Reference)

☞ Encyclopaedias

☞ FAQs (e.g., www.dynamicmedical.com: GPs)
Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion
Taxonomies of Medical Questions

- Clinical questions by “family doctors”
- Clinical questions by (practicing) medical students
A taxonomy of generic clinical questions

- 1396 questions collected from 152 family doctors (plus 14 general internists and 6 general paediatricians)

- Goal-oriented classification: *diagnosis, treatment, management, epidemiology, non-clinical*

- Four-level (max) hierarchy

- A different perspective: contrasts with form-oriented classifications (*Graesser, etc.*), although sometimes congruent at inner levels (*e.g.*, *name finding*)

A few generic question types account for a large number of questions

*(Ely et al., BMJ 2000)*
**Taxonomy of Generic Question Types (excerpts)**

**Diagnosis** cause/interpretation of clinical finding

\[
\text{What is the cause of } \left\{ \begin{array}{l}
\text{symptom} \\
\text{clinical finding} \\
\text{test finding}
\end{array} \right\} \text{?} \quad (21.3\%)
\]

- What is the differential diagnosis of symptom X?
- What is the likelihood that symptom X is coming from condition Y?
- At what level does physical finding X become clinically important?
- How should I use test finding X in my decision?
**Taxonomy of Generic Question Types (excerpts)**

**Diagnosis**

Test (lab test, skin test, biopsy, imaging, element of physical exam, etc.); indications / efficacy, accuracy, timing, etc.

*What is the best test in situation Y?*, questions related to test properties (11.8%)

- *Is test X indicated in situation Y?*
- *How do I diagnose condition Y?*
- *How do I distinguish between conditions Y₁, Y₂, ... Yₙ?*
- *How good is test X in situation Y?*
- *When (timing) should I do test X?*
**Taxonomy of Generic Question Types (excerpts)**

**Treatment** Drug prescribing; how to prescribe; dosage, timing, undifferentiated

*Dose of drug X?* (9.3%)

- What is the dose of drug X?
- Should I change the dose of drug X (in situation Y)?
- When should I \{ start, stop \} drug X?
Taxonomy of Generic Question Types (excerpts)

**Treatment** Drug prescribing; efficacy / indications / drug of choice / treatment

- *Is drug X indicated in situation Y?* (10.7%)
- *How effective is drug X for condition Y?*
- *What is the drug of choice for condition Y?*
- *Is drug X₁ better than drug X₂, X₃, . . . Xₙ for condition Y?*
Taxonomy of Generic Question Types (excerpts)

**Treatment**  Not only drug prescribing; efficacy/indications; treatment

- How should I treat \( \{ \text{finding} \} \) \( \text{condition} \) \( Y \) (given situation \( Z \))? (5.9%)

- Should I use \( \{ \text{treatment} \} \) \( \text{procedure} \) \( X \) for \( \{ \text{condition} \} \) \( \text{finding} \) \( Y \)?

- What is the efficacy of \( \{ \text{treatment} \} \) \( \text{procedure} \) \( X \) (for condition \( Y \))?

- Is there any treatment for condition \( Y \)?

- At what level of severity of condition \( Y \) is treatment indicated?
**Management**  condition / finding

\[
\text{How should I manage } \begin{cases} \text{condition} \\ \text{finding} \\ \text{situation} \end{cases} Y? (4.8%)
\]

- What management options are there in situation Y?
- How aggressive/conservative should I be in situation Y?

Total: 63.8%
Clinical Questions by Medical Students

(Jacquemart & Zweigenbaum, Stud Health Technol Inform 2003)

- 100 questions collected from medical students (oral pathology)

- Quels signes caractérisent un lichen plan?
  *Which signs characterize a lichen planus?*

- Quelle mesure diagnostique une adénopathie?
  *Which test diagnoses an adenopathy?*

- Doit-on traiter les atteintes buccales du lichen plan?
  *Must one treat oral lesions of lichen planus?*
Classification by form and content:

- patterns [A]–(R)–[B], with A and B medical concept types (pathology, treatment, etc.) and R a UMLS relation (causes, diagnoses, etc.)
  
  e.g., [Observation]–(diagnoses)–[Pathology]
  
  *(Which signs characterize a lichen planus?)*

- focus on A or B (68%)

  *The Reed-Sternberg cell evokes which disease?*

  [A]–(diagnoses)–[which B]

Cf. *(Lin J., 2001)*’s ternary expressions?
Clinical Questions by Medical Students

- focus on whole triple (17%)
  
  Does diabetes contraindicate vasoconstrictors?
  Is it true that [A]–(complicates)–[B]?

- explanation on whole triple (5%)

  Pourquoi séparer un germe de dent de sagesse inclus?
  why [A]–(treats)–[B]

Total: 80% triple-based questions
Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion
Terminological Resources: Identify Named Entities

An extended definition of named entities: specific names useful in the domain

- Person names, company names, location names, etc.
  Lists (gazetteer, etc.), learning methods

- Biomedical names:
  - drug: codeine; aspirin; acetyl salicylate
  - finding: decreased capillary fragility
  - test finding: Abnormal skin pH
  - symptom: dyspnea; nausea; Homans’ sign
  - diagnosis: chronic obstructive lung disease; COLD
  - test: radioimmunoassay; acoustic rhinometry
  - treatment: hemodialysis; cesarean section
  - gene: c-Ha-ras

Use biomedical terminologies to identify biomedical names
Hierarchical navigation: generic / specific terms

- **General English**: WordNet

- **Biomedical domain**:

<table>
<thead>
<tr>
<th>Term</th>
<th>Relation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>infection</td>
<td>is-a</td>
<td>disease</td>
</tr>
<tr>
<td>abscess</td>
<td>is-a</td>
<td>infection</td>
</tr>
<tr>
<td>salivary gland</td>
<td>is-a</td>
<td>digestive organ</td>
</tr>
<tr>
<td>bile</td>
<td>part-of</td>
<td>biliary system</td>
</tr>
<tr>
<td>bile duct</td>
<td>part-of</td>
<td>biliary system</td>
</tr>
<tr>
<td>nucleus</td>
<td>part-of</td>
<td>cell</td>
</tr>
<tr>
<td>chromatin</td>
<td>part-of</td>
<td>cell</td>
</tr>
</tbody>
</table>

Use **biomedical terminologies** to identify hierarchical relations
Biomedical domain: more than 100 terminological products!
Thesauri, classifications, nomenclatures...
Medical Terminologies: General

(Cimino, IMIA Yearbook 1996)

- **ICD**: International Classification of Diseases (WHO)
  ICD-9 (1977), ICD-9-CM, ICD-10 (1993), ...: Diagnoses

- **MeSH**: Medical Subject Headings (NLM)
  Yearly updates; Indexes the Medline bibliographic database

  Clinical orientation

- **ICPC**: International Classification of Primary Care
  General Practitioners
Medical Terminologies: Specific

Specific to subdomain, to medical specialty, to country, to user group

- **UWDA**: University of Washington Digital Anatomist (2001)
  Anatomy

- **SNOMED Microglossary** for Pathology
  Anatomo-pathology

- Minimal Standard Terminology for **Digestive Endoscopy**

- **DSM-IV**: Diagnostic and Statistical Manual of Mental Disorders
  Psychiatry

- **CPT4**: Physicians’ Current Procedural Terminology
  Procedures

- **MedDRA**: Medical Dictionary for Regulatory Activities
  Drugs

PZweigenbaum, AP Hôp Paris & ERM 202 INSERM EA CL 2003 NLP4QA
Medical Terminologies: Structure

Strongly structured terminologies (generally)

✗ Synonym terms (often)

✗ Hierarchical structure (most)
  ➢ is-a
  ➢ part-of

✗ Other relations (sometimes)
The **Unified Medical Language System**

(Lindberg et al., Meth Inf Med 1993; McCray & Nelson, Meth Inf Med 1995)

The UMLS project is a long-term NLM research and development effort designed to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources. (UMLS Documentation, 2002) [umlsinfo.nlm.nih.gov](http://umlsinfo.nlm.nih.gov)

- Metathesaurus
- Semantic Network
- Specialist Lexicon

Distributed free of charge (sign convention with NLM); individual vocabulary sources may have usage restrictions
The goal is to make it easier to develop systems that link information from patient record systems, bibliographic databases, factual databases, expert systems, etc. The UMLS Knowledge Sources can also facilitate the development of data creation and indexing applications. (UMLS Documentation, 2002)

- Houses the contents (terms, concepts and relations) of 100 medical terminologies (2003AA)
  - 800,000 concepts
  - 2,100,000 strings
- Identifies common concepts
  Maps concepts of individual terminologies into unique Metathesaurus space
What the Metathesaurus Is

- A repository of biomedical terms
  - Addresses the issue of biomedical named entities

- A repository of families of “synonymous” terms
  - Addresses the issue of terminological variation

- A huge network of interconnected concepts
  - Addresses the issue of hierarchical navigation

- Not a terminology, but several overlayed terminologies

- Not an ontology: multiplicity of non-homogeneous structures
The UMLS Semantic Network

- Adds its own domain structuring over the Metathesaurus
  - 134 semantic types (tree)
  - Hierarchy of 54 relations
  - Relation “signatures”
- Each Metathesaurus concept has one or more semantic types
- Most Metathesaurus relations are instances of Semantic Network relations

Unified, general categorization of medical entities
### The UMLS Semantic Network: Examples

<table>
<thead>
<tr>
<th>(Ely et al.)</th>
<th>UMLS</th>
<th>Example Metathesaurus concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>finding</td>
<td>Finding</td>
<td>Birth History; Downward displacement of diaphragm</td>
</tr>
<tr>
<td>test finding</td>
<td>Laboratory or Test Result</td>
<td>Forced Expiratory Volume; Abnormal skin pH</td>
</tr>
<tr>
<td>symptom</td>
<td>Sign or Symptom</td>
<td>Dyspnea; Nausea; Pain; Pallor; Homans’ sign</td>
</tr>
<tr>
<td>test</td>
<td>Laboratory Procedure</td>
<td>Blood Protein Electrophoresis; Radioimmunoassay</td>
</tr>
<tr>
<td>test</td>
<td>Diagnostic Procedure</td>
<td>Heart Auscultation; Tomography, X-Ray Computed</td>
</tr>
<tr>
<td>treatment</td>
<td>Therapeutic or Preventive Procedure</td>
<td>Cesarean Section; Hemodialysis; Vaccine Therapy</td>
</tr>
<tr>
<td>drug</td>
<td>Pharmacologic Substance</td>
<td>Antiemetics; Cardiovascular Agents; Codeine</td>
</tr>
</tbody>
</table>
**UMLS and Language: the Specialist Lexicon**

- A medical lexicon, with general applicability
- 163,000 entries, 268,000 inflected forms
- Morphological and syntactic information
- Lexical tools: tokenize, normalize string, compute inflected forms, derived words, etc.
  
  *eye, oculus, ocular*

  - Addresses the issue of terminological variation

- A specialized indexer’s toolkit
- No usage restrictions
Biomedical Term Indexing with the UMLS

(Aronson et al., AMIA 2000, AMIA 2001)

Identify UMLS concepts in a document:

- ako Named Entity Recognition
- MetaMap program
- Cumulates the lexical variants of the Specialist Lexicon with the terminological variants of the Metathesaurus
- Used in the NLM Indexing Initiative

Can be obtained free of charge; no usage restrictions
Caveat: UMLS Non-English Content (terms)

e.g., French content?

Metathesaurus: 25,000 concepts (2002)
  i.e., 3% of Metathesaurus!
  - MeSH (29,173 strings)
  - WHO adverse reaction terminology (3,717)
  - digestive endoscopy (1,833)
  - primary care (723)

Currently, in French, UMLS terms ↔ MeSH terms

Similar situation for Dutch, Finnish, Italian, Portuguese, Russian, and Spanish; German slightly better off (ICD-10).

➡️ project VUMeF: add more French terms
Caveat: UMLS Non-English Content (lexicon)

Non-English medical lexicons?

- DSL: German Specialist Lexicon (inflection, derivation, compounding) \((\text{Weske-Heck et al., AMIA 2002})\)

- project UMLF: French Specialist Lexicon in preparation (inflection, derivation) \((\text{Zweigenbaum et al., MIE 2003})\)
In Need of an Ontology?

What is an ontology?

➢ The UMLS is not an ontology
  (does not support formal reasoning)

➢ GALEN ontology for medicine

➢ Gene Ontology for genomics
The GALEN Ontology

A General Architecture for Language and Nomenclatures *(Rector et al., MEDINFO 1992)*

➢ Knowledge representation formalism (GRAIL, description logic)
➢ Define new concepts through composition
➢ Automatic classification
➢ Words or terms associated “externally” to concepts

➠ Sound, formal concept definitions;
  more complex than simple thesaurus
➠ Generative, potentially open;
  current coverage lower than UMLS

Concept definitions downloadable from www.opengalen.org
Gene product, molecular function, biological process, cellular component

Terms, synonyms

Multiple hierarchies, explicit hierarchical relations:

- is-a
- part-of

Controlled vocabulary

Downloadable from www.geneontology.org
Task: Which Questions for Whom

Task: Specificities of Medical Language Sources

Resources: Health Information Sources

Resources: Types of Questions

Resources: Linguistic and Terminological Bases

Conclusion
**Synthesis: Challenges and Resources**

- **Challenges** for question-answering systems

- **Resources to address these challenges:**
  - quality-controlled health information gateways offer a thorough indexing of *trustworthy biomedical sources*
  - *taxonomies of question types* rank and categorize interesting questions, taking into account their frequency of occurrence
  - biomedical lexicons, terminologies and ontologies are there to help manage *domain-specific terms and concepts*
**Perspective: Evaluation of Biomedical QA**

- EQueR evaluation of French QA systems *(Grau, 2003)* in the context of the Technolangue EVALDA initiative for evaluating NLP systems (ELRA)

- (French) Medical QA track: participation welcome!
Acknowledgements

Many thanks to

✗ Aurel Craciun
✗ Tiphaine Dalmas
✗ Stéfan Darmoni
✗ Pierre Jacquemart