Annotation Scheme of *HwaMei* Clinical Corpus

A Preprint

Enwei Zhu
Ningbo Institute of Life and Health Industry
University of Chinese Academy of Sciences
Ningbo 315016, Zhejiang Province, P. R. China
zhuenwei@ucas.ac.cn

August 5, 2021

Abstract

This is the annotation scheme of *HwaMei* clinical corpus. The scheme covers three aspects of medical information, i.e., medical entity, relation and attribute. This technical report in detail describes the annotation principles, formats, definitions along with extensive examples for each type of medical entity, relation and attribute. Our annotation scheme consists of nine entity types (18 sub types), ten relation types and ten attribute types. The medical relation and attribute types are defined with their applicability to corresponding entity types; and the relations are, in particular, elegantly formulated between abstract roles that medical entities may play.

Keywords Information Extraction · Annotation Scheme · Clinical Text

1 Introduction

This technical report describes the medical information annotation scheme of *HwaMei* clinical corpus in detail, which serves as a supplementary material of our working paper [Zhu et al., 2021]. The annotation scheme is based on comprehensive reviews of previous studies [Uzuner et al., 2011, Roberts et al., 2009, He et al., 2017, Campillos et al., 2018] and multiple-round consultations with experienced medical experts.

A comprehensive information annotation scheme typically comprises three aspects, i.e., medical entity, relation and attribute. Medical entity is the most basic information element. Early studies focused on a single entity type such as medical problem [Meystre and Haug, 2006, Ogren et al., 2008] or medication [Uzuner et al., 2010], while a relatively comprehensive scheme would include at least disease, symptom, test, treatment and drug [Roberts et al., 2009, Uzuner et al., 2011]. Medical relation types are defined between each pair of entity types. This typically yields a large number of relation types [Campillos et al., 2018, Guan et al., 2020], which may easily cause ambiguity between different types and thus confuse the annotators. Medical attribute is also called assertion, which indicates some special statuses of an entity. The most basic attribute types are negation and uncertainty [Uzuner et al., 2011]. These annotation schemes are either not comprehensive, or short of consistency and cooperativeness to some extent; more specifically, the entity, relation and attribute types are typically separately defined, with limited illustration of the compatibility between them. Additionally, the corpus scales are relatively small, which may not be sufficient to build a reliable medical information extraction system.

Our annotation scheme is designed to be simple but comprehensive. It consists of nine entity types (18 sub types), ten relation types and ten attribute types. The relation and attribute types are defined with their applicability to corresponding entity types, which makes them of higher integrity. In particular, the medical relations are elegantly formulated between three abstract roles (Status, Information and Intervention) that medical entities may play. The annotation scheme generally follows four principles:

Collectively exhaustive The definitions of the medical entities, relations and attributes should be as comprehensive as possible, so that the extracted elements are capable of describing and representing the medical information in clinical texts. Ideally, all the medical information in the original clinical texts can be retrieved via the extracted medical elements.

Mutually exclusive The definitions of medical entities, relations and attributes should be as clear as possible at the boundaries between different types. It should be avoided that an entity mention (or an entity pair) can be legally assigned to multiple entity (or relation) types, which will confuse the annotators and lead to inconsistent annotation results.

Semantics are contextualized The clinical text should be interpreted along with its contextual information. Even a same word or phrase may express completely different semantics and thus require different annotations, given the contexts are different. In particular, the annotator should not mechanically map words and phrases to fixed annotation results, just like looking up a fixed table. For example, any abbreviation or descriptive phrase should be annotated as the corresponding medical element, provided that the context is clear [Uzuner et al., 2011].

Literalness is prior to external knowledge The annotations should be firstly in accordance to the text literalness, and secondly the annotator's medical knowledge background. As an extreme (and infrequent) case, the text may be vague or just have some mistakes, and thus conflict with the annotator's medical knowledge; in this case, the annotator should follow the literalness.[1] Otherwise, if the text implicitly mentions a medical element, or in other words, it can be inferred according to the annotator's medical knowledge, it should be annotated only when the knowledge is a widely recognized consensus (e.g., the normal body temperature range) [Roberts et al., 2009].

The rest of this report provides detailed annotation guidelines with extensive examples for medical entity (Section 2), relation (Section 3) and attribute (Section 4).

2 Medical Entity

Given an input text as a sequence of tokens $x = x_1, x_2, \ldots, x_T$ of length T, the objective output of entity recognition is a set y^{Ent}, of which each element is a triplet of entity type, start position, and end position:

$$y^{Ent} = \{(type_i^{Ent}, start_i, end_i) \mid type_i^{Ent} \in S^{Ent}, 0 \leq start_i, end_i < T, i = 1, 2, \ldots, N\} \quad (1)$$

where S^{Ent} is the set of entity types, N is the number of entities in the text. Correspondingly, the annotation of a entity should specify its type, start and end positions.

We formulate nine types of entities: Disease, Symptom, Test, Treatment, Drug, Body, Personal History, Equipment and Department. Each entity type may consist of multiple sub types, which further characterize the nature of entities.

2.1 Disease

Disease refers to a *condition* or *status* of the patient which impairs normal functioning,[2] and is typically manifested in distinguishing symptoms and signs. It is, in most cases, interchangeable with the concept of diagnosis in clinical practice. This entity type consists of three sub types: Disease or Syndrome, Injury or Poisoning and Organ Damage.

Disease or Syndrome Disease or Syndrome is a sub type which emphasizes the nature that these diseases are mainly developed by internal causes instead of external ones.

- ……否认心脏病、糖尿病史……(. . . denies heart disease or diabetes history . . .)
  ⇒ 心脏病 (heart disease), 糖尿病 (diabetes)

[1] The corpus is ultimately used for building a system to extract medical information from real clinical texts. Hence, we have to respect the text literalness, assuming that that the input clinical texts are generally reliable with minor mistakes, and the algorithms should ideally be able to detect and eliminate the mistakes.

[2] The type (e.g., diabetes type I, type II) and stage (e.g., early, late liver cancer) information should also be included in the span of a Disease entity.

- 初步诊断：类风湿性关节炎……(Initial diagnosis: rheumatoid arthritis . . .)
  ⇒ 类风湿性关节炎 (rheumatoid arthritis)
- 患者至我院要求手术治疗，诊断为 “肝恶性肿瘤” (The patient came to our hospital for surgery and was diagnosed as “liver malignant tumor”)
  ⇒ 肝恶性肿瘤 (liver malignant tumor)

Injury or Poisoning Injury or Poisoning is a sub type which emphasizes the nature that these diseases are mainly developed by external causes instead of internal ones.

- 初步诊断：右膝关节软骨撕裂……(Initial diagnosis: torn cartilage of the right knee . . .)
  ⇒ 右膝关节软骨撕裂 (torn cartilage of the right knee)
- 因头部外伤入院……(. . . admitted to hospital due to head trauma . . .)
  ⇒ 头部外伤 (head trauma)

Organ Damage Organ Damage is a sub type which refers to the damage, dysfunction, disturbance or insufficiency of an organ or body part. It describes the abnormal status itself, but implies limited clues of the causes. It typically appears in clinical analysis.

- 有肝功能异常、肾损害……(. . . have abnormal liver function, kidney damage . . .)
  ⇒ 肝功能异常 (abnormal liver function), 肾损害 (kidney damage)
- 胸部 CT 提示 “气管狭窄” ……(Chest CT suggests “tracheal stenosis” . . .)
  ⇒ 气管狭窄 (tracheal stenosis)
- 术中无血管神经损伤等并发症 (No complications such as injury of blood vessels and nerves during the surgery)
  ⇒ 血管神经损伤 (injury of blood vessels and nerves)

2.2 Symptom

Symptom refers to a *phenomenon* or *manifest* which is typically caused by, and in turn indicates the existence of diseases. It consists of two sub types: Self-Reported Abnormality and Abnormal Test Result.

Self-Reported Abnormality Self-Reported Abnormality refers to the patient's subjective discomfort, which is typically self-reported rather than diagnosed by the physician.

- 经长时间直立行走后，患者出现右侧腹股沟坠胀感……(After long-time walking, the patient felt bulging in the right groin. . .)
  ⇒ 坠胀感 (bulging)
- 患者于 7 年前无明显诱因下在家中出现乏力，以活动后为甚……(The patient developed fatigue at home without obvious incentives 7 years ago, especially after activities . . .)
  ⇒ 乏力 (fatigue)
- 因 “腹痛 3 天” 入院 (Admitted to the hospital with “abdominal pain for 3 days”)
  ⇒ 腹痛 (abdominal pain)
- 小便稍红，大便正常……(Urine is slightly red, stool is normal . . .)
  ⇒ 小便稍红 (Urine is slightly red)

Abnormal Test Result Abnormal Test Result refers to abnormal evidence found by physical examination, laboratory or imaging tests. A crucial detail is that an entity of Abnormal Test Result must be an *abnormal* result (i.e., deviated from a healthy status); otherwise, an entity referring to a *normal* result should be annotated as the Test Result type (defined in Subsection 2.3). In practice, if the context itself implies the abnormality, the entity should be inferred as *abnormal*; if the deduction relies on the annotator's medical knowledge background, the entity should be inferred as *abnormal* only when the knowledge is a widely recognized consensus (e.g., the normal body temperature range).

- ……脾显著肿大, 腹腔大量积液……(. . . the spleen is significantly enlarged, and the abdominal cavity has a large amount of fluid . . .)
  ⇒ 脾显著肿大 (the spleen is significantly enlarged), 腹腔大量积液 (the abdominal cavity has a large amount of fluid)

- 腹部 CT 平扫 + 增强：胃角处溃疡，符合胃癌表现……(Abdominal CT plain scan + enhanced: Ulcer at the corner of the stomach, in line with gastric cancer . . .)
  ⇒ 胃角处溃疡 (Ulcer at the corner of the stomach)
- 因肝酶升高明显，患者昨至医院就诊，查 "血常规 +CRP" ……(Because liver enzymes were significantly elevated, the patient went to the hospital yesterday to check "blood routine + CRP" . . .)
  ⇒ 肝酶升高明显 (liver enzymes were significantly elevated)
- 下腹部可见陈旧性手术疤痕，四肢肌力基本正常，肌张力正常 (Old surgical scars can be seen in the lower abdomen, the muscle strength of the limbs is basically normal, and the muscle tone is normal)
  ⇒ 下腹部可见陈旧性手术疤痕 (Old surgical scars can be seen in the lower abdomen)

2.3 Test

Test refers to a process for searching clues of a disease, such as physical examination, laboratory and imaging tests. An entity of Test is typically not aimed to cure diseases or relieve symptoms. It consists of two sub types: Test Process and Test Result.

Test Process Test Process refers to a test service provided to patients. An entity of Test Process can be a specific test item, or a group of test items.

- 腹部 CT 平扫 + 增强：胃角处溃疡，符合胃癌表现……(Abdominal CT plain scan + enhanced: Ulcer at the corner of the stomach, in line with gastric cancer . . .)
  ⇒ 腹部 CT 平扫 + 增强 (Abdominal CT plain scan + enhanced)
- 因肝酶升高明显，患者昨至医院就诊，查 "血常规 +CRP" ……(Because liver enzymes were significantly elevated, the patient went to the hospital yesterday to check "blood routine + CRP" . . .)
  ⇒ 血常规 (blood routine), CRP (CRP)
- PPD 试验和 X 线检查有助于鉴别……(PPD test and X-ray examination helps to identify . . .)
  ⇒PPD 试验 (PPD test), X 线检查 (X-ray examination)
- 拟明日复查肝肾功能、电解质，继续观察病情 (Plan to re-test liver and kidney functions, electrolytes tomorrow, and continue to observe the patient's conditions)
  ⇒ 肝肾功能 (liver and kidney functions), 电解质 (electrolytes)

Test Result Test Result refers to a test item along with its result (including the number and unit, if applicable). Note that an entity of Test Result must be a *normal* result; otherwise, it should be annotated as the Abnormal Test Result type. See Subsection 2.2 for more details.

- 生命体征：脉搏 52 次/分，血压 125/86mmHg，呼吸 19 次/分，氧饱和度 100% (Vital signs: pulse 52 beats/min, blood pressure 125/86mmHg, breathing 19 beats/min, oxygen saturation 100%)
  ⇒ 脉搏 52 次/分 (pulse 52 beats/min), 血压 125/86mmHg (blood pressure 125/86mmHg), 呼吸 19 次/分 (breathing 19 beats/min), 氧饱和度 100% (oxygen saturation 100%)
- 肝肾功能：丙氨酸氨基转移酶 806U/L，天门冬氨酸氨基转移酶 809U/L……(Liver and kidney functions: alanine aminotransferase 806U/L, aspartate aminotransferase 809U/L . . .)
  ⇒ 丙氨酸氨基转移酶 806U/L (alanine aminotransferase 806U/L), 天门冬氨酸氨基转移酶 809U/L (aspartate aminotransferase 809U/L)

2.4 Treatment

Treatment refers to a process aimed to cure diseases or relieve symptoms, such as surgeries, chemotherapy and radiotherapy. It does not include drugs, which should be annotated as the Drug type (defined in Subsection 2.5). It consists of four sub types: Treatment, Operation, Prevention and Care.

Treatment Treatment refers to a complete treatment process, such as surgeries, chemotherapy and radiotherapy. Treatment focuses on the diseases or symptoms currently present on patient; on the other hand, those aimed to prevent potential diseases or symptoms in the future should be annotated as the Prevention type.

- 排除手术禁忌后，择期行腹股沟斜疝无张力修补术 (After eliminating surgical contraindications, plan for tension-free repair of indirect inguinal hernia)
  ⇒ 腹股沟斜疝无张力修补术 (tension-free repair of indirect inguinal hernia)

- 患者已肿瘤晚期，术前化疗可改善患者预后，延长生命 (The patient has advanced tumor, preoperative chemotherapy can improve the prognosis and prolong the life of the patient)
  ⇒ 化疗 (chemotherapy)
- 患方拒绝抗血管治疗、靶向治疗、免疫治疗、化疗等全身治疗方案 (The patient refused anti-vascular therapy, targeted therapy, immunotherapy, chemotherapy and other systemic treatment options)
  ⇒ 血管治疗 (anti-vascular therapy), 靶向治疗 (targeted therapy), 免疫治疗 (immunotherapy), 化疗 (chemotherapy)
- 患者左膝🄵窝部曾行切开排脓术 (The patient's left knee popliteal fossa had undergone incision and drainage)
  ⇒ 左膝🄵窝部曾行切开排脓术 (left knee popliteal fossa had undergone incision and drainage)

Operation Operation refers to a specific step of treatment. It does not form a complete treatment process, and is not able to independently achieve the treatment purpose. Operation entities typically appear in text that describes surgery procedures.

- 1% 利多卡因局麻，以 Seldinger's 法穿刺右桡动脉顺利置入 6F 动脉鞘……(1% lidocaine was used for local anesthesia, and the right radial artery was punctured by Seldinger's method and the 6F arterial sheath was smoothly inserted . . .)
  ⇒ 局麻 (local anesthesia), 以 Seldinger's 法穿刺右桡动脉 (the right radial artery was punctured by Seldinger's method)
- 结束手术，退管拔鞘，手法压迫止血 (End the surgery, withdraw the tube and remove the sheath, and use compression to stop bleeding)
  ⇒ 退管拔鞘 (withdraw the tube and remove the sheath), 手法压迫止血 (use compression to stop bleeding)

Prevention Prevention refers to a treatment process aimed to prevent potential diseases or symptoms in the future.

- "复方磺胺甲🄵唑" 预防卡肺 (Use "compound sulfamethoxazole" to prevent cardiac pneumonia)
  ⇒ 预防卡肺 (prevent cardiac pneumonia)
- 进行抗凝血治疗，预防休克 (Take anticoagulant and anti-shock therapies)
  ⇒ 抗凝血 (anticoagulant), 预防休克 (anti-shock)
- 注入地尔硫卓 3mg 预防冠脉痉挛……(Inject diltiazem 3mg to prevent coronary artery spasm . . .)
  ⇒ 预防冠脉痉挛 (prevent coronary artery spasm)

Care Care refers to a nursing care process.

- 继续流质饮食，适当营养支持治疗 (Continue the liquid diet and provide appropriate nutritional support and treatment)
  ⇒ 流质饮食 (liquid diet)
- 予 I 级护理，普食 (Give level I care, general food)
  ⇒I 级护理 (level I care), 普食 (general food)

2.5 Drug

Drug refers to a medicine for curing diseases or relieving symptoms. It consists of two sub types: Drug and Drug Dose.

Drug Drug refers to the drug name with its specification.

- 1% 利多卡因局麻，以 Seldinger's 法穿刺右桡动脉顺利置入 6F 动脉鞘……(1% lidocaine was used for local anesthesia, and the right radial artery was punctured by Seldinger's method and the 6F arterial sheath was smoothly inserted . . .)
  ⇒1% 利多卡因 (1% lidocaine)
- 出院后口服护肝降酶药物，注意复查肝功能电解质 (After discharge from the hospital, take orally liver-protecting and enzyme-lowering drugs, pay attention to rechecking liver function and electrolytes)
  ⇒ 护肝降酶药物 (liver-protecting and enzyme-lowering drugs)

- 出院药物医嘱：复方甘草酸苷片 2 片一次，一日 3 次口服 (Discharge medical order: compound glycyrrhizin tablets 2 tablets once, 3 times a day, orally)
  ⇒ 复方甘草酸苷片 (compound glycyrrhizin tablets)

Drug Dose Drug Dose refers to the drug quantity (including the number and unit, if applicable).

- 出院药物医嘱：复方甘草酸苷片 2 片一次，一日 3 次口服 (Discharge medical order: compound glycyrrhizin tablets 2 tablets once, 3 times a day, orally)
  ⇒2 片一次，一日 3 次 (2 tablets once, 3 times a day)

2.6 Body

Body consists of two sub types: Body Part and Body Matter. It is very common that a Body entity overlaps, or more particularly, is contained in a entity of other types. In that case, we would annotate the other entity instead of the Body entity. In other words, Body entities are of lower priority in annotation.[3]

Body Part Body Part refers to an organ, tissue or location of a human body.

- ……常位于两上肺尖后段及下叶背段……(. . . often located in the posterior part of the upper lung apex and the dorsal part of lower lobe . . .)
  ⇒ 两上肺尖后段 (the posterior part of the upper lung apex), 下叶背段 (the dorsal part of lower lobe)
- 腋后线第 8 肋间置入艾贝尔引流管一根 (An Abel drainage tube is inserted into the 8th intercostal space of the posterior axillary line)
  ⇒ 腋后线第 8 肋间 (the 8th intercostal space of the posterior axillary line)

Body Matter Body Matter includes body fluid, secretion, urine or excrement.

- 患者出现恶心呕吐，为胃内容物，无发热畏寒、无胸闷 (The patient has nausea and vomiting, stomach contents, no fever or chills, no chest tightness)
  ⇒ 胃内容物 (stomach contents)

2.7 Personal history

Personal History refers to the patient's history activities that are related to medical problems.

- 患者否认吸烟、饮酒史，否认吸毒史 (The patient denies the history of smoking and drinking, and denies the history of drug use)
  ⇒ 吸烟 (history of smoking), 饮酒史 (drinking), 吸毒史 (history of drug use)

2.8 Equipment

Equipment includes medical equipment, instruments and devices. Similar to Body ,Equipment entities are of lower priority in annotation.

- 固定起搏电极，调整起搏参数及电压 (Fix the pacing electrode, adjust the pacing parameters and voltage)
  ⇒ 起搏电极 (pacing electrode)
- 用电刀切断两侧直肠侧韧带 (Use an electric knife to cut off the collateral ligaments on both sides)
  ⇒ 电刀 (electric knife)

2.9 Department

Department refers to the name of a hospital department.

- 患者几个月前现痤疮样皮疹，在皮肤科就诊 (The patient had an acne-like rash a few months ago, and went to the dermatology department)
  ⇒ 皮肤科 (dermatology department)

[3]This principle is consistent with previous literature. In addition, many medical terminologies also include body part/matter, e.g., heart disease, chest CT.

Table 1: Medical entity types

Entity type	Sub types	Annotation priority
Disease	Disease or Syndrome, Injury or Poisoning, Organ Damage	High
Symptom	Self-Reported Abnormality, Abnormal Test Result	High
Test	Test Process, Test Result	High
Treatment	Treatment, Operation, Prevention, Care	High
Drug	Drug, drug Dose	High
Body	Body Part, Body Matter	Low
Personal history		Low
Equipment		Low
Department		Low

2.10 Summary

Table 1 presents the nine medical entity types with corresponding sub types.

Flat entity scheme and annotation priority We employ a *flat* scheme for entity annotation, which means that no nested or overlapping entities are allowed. This is a trade-off result; otherwise, there would be too many trivial and redundant annotations. For example, it is a common case that a Disease or Symptom entity includes a Body Part entity in its name. The Disease or Symptom entity is a whole concept, which has already indicated the corresponding body part, so explicitly annotating the nested Body Part entity provides limited incremental information.

To this end, we should design which entities should be preferentially annotated in potential clashing cases. In general, the former five types (Disease, Symptom, Test, Treatment and Drug) are regarded to be more important than the later four types (Body, Personal History, Equipment and Department); hence, the former five entity types are of relatively high priority in annotation. For example, in text "the patient has heart disease ...", a Body Part entity is inside a Disease entity, then the prior Disease entity "heart disease" should be annotated while the Body Part entity "heart" should be left unannotated.

Interpreting abbreviations It is common that a medical entity appear as an abbreviation in clinical text. These abbreviations should be interpreted and annotated as the corresponding original entities. For example, in text "the patient should check the kidney function and blood sugar ... ", the span "blood sugar" does refer to a medical test based on the context, although the span itself may not necessarily refer to it. In this case, we would still treat the text span as a valid mention of the entity.

3 Medical Relation

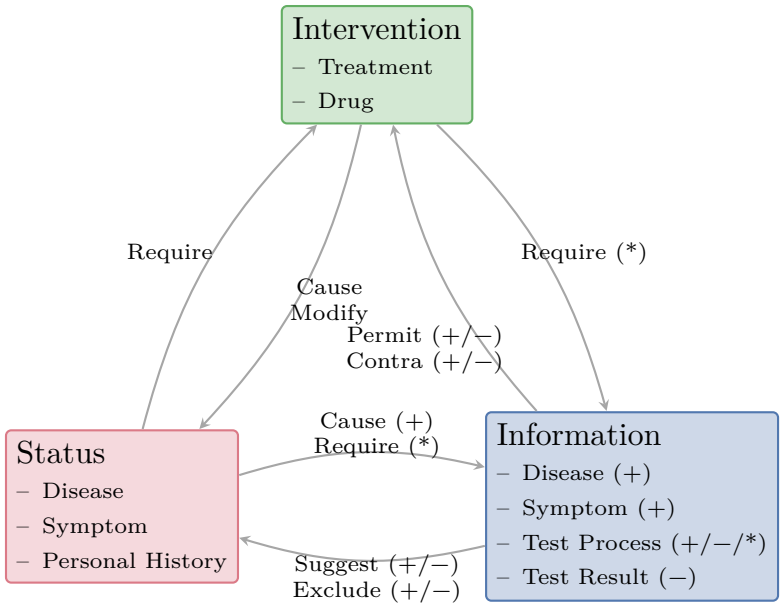
Given tokens x and entities y^{Ent}, the objective output of relation extraction is a set y^{Rel}, of which each element is a triplet of relation type, head entity and tail entity:

$$y^{Rel} = \{(type_j^{Rel}, head_j, tail_j) \mid type_j^{Rel} \in S^{Rel}, head_j, tail_j \in y^{Ent}, j = 1, 2, \ldots, M\} \tag{2}$$

where S^{Rel} is the set of relation types, M is the number of relations between the entities. Correspondingly, the annotation of a relation should specify its relation type, head and tail entities.

Previous studies typically enumerate every pair of entity types and define multiple relation types between each pair [Uzuner et al., 2011, He et al., 2017, Campillos et al., 2018]. This approach may easily lead to either omissions or redundancies of the relation type coverage, especially when the entity types are defined of a fine granularity. An entity of a specific type may play different roles relative to another entity, which results in many possible relation types. Based on multiple-round in-depth discussion with medical experts, we formulated medical relations between three essential "entity roles", namely "Status", "Information" and "Intervention".

- **Status**: The patient's essential, underlying health conditions, which may cause certain relatively non-essential appearances. A Status role is typically played by a Disease, Symptom or Personal History entity.
- **Information**: The patient's non-essential appearances, which is typically caused by relatively essential underlying health conditions. It can be further divided into three sub roles.



Notes: (1) This figure presents the potential medical relations between the Status, Information and Intervention roles. Each role can be played by a entity of specific types. The positive/negative/star sign marked on the right of an entity type under the Information role indicates that the entity type can play Positive/Negative/Unknown Information. The positive/negative/star sign marked on the right of a relation type indicates that the relation applies to Positive/Negative/Unknown information. (2) If an entity type is listed, the corresponding relations apply to all its sub types.

Figure 1: Medical relations between entities as Status, Information and Intervention

  - **Positive Information**: An abnormal appearance, typically played by a Disease, Symptom (including Self-Reported Abnormality and Abnormal Test Result) or Test Process entity.
  - **Negative Information**: A normal appearance, typically played by a Test Result or Test Process entity.
  - **Unknown Information**: An appearance unknown whether it is normal or not, typically played by a Test Process entity.
- **Intervention**: The physician's measures aimed to improve the patient's health conditions. An Intervention role is typically played by a Treatment or Drug entity.

Figure 1 presents the potential medical relations between the three roles. We argue that most essential relation types have been covered in this framework.

3.1 Status vs. Information

Status–Cause–Information The Status entity is one of the potential reasons or triggers of the Information entity, where the Information is supposed to be positive. In this relation, the Status entity is the relative essence, while the Information is the relative appearance. A typical example is that a Disease entity causes a Symptom entity.

- 高血压容易引起头晕、头痛等症状 (High blood pressure can easily cause dizziness, headaches and other symptoms)
  ⇒ 高血压 (High blood pressure) – 头晕 (dizziness); 高血压 (High blood pressure) – 头痛 (headaches)
- 下肢深静脉血栓形成: 常表现为一侧下肢突发肿胀, 疼痛, 皮肤紫色……(Deep vein thrombosis of the lower extremities: often manifested as sudden swelling and pain on one side of the lower extremity, and the skin becomes purple . . .)
  ⇒ 下肢深静脉血栓 (Deep vein thrombosis of the lower extremities) – 一侧下肢突发肿胀, 疼痛 (sudden swelling and pain on one side of the lower extremity); 下肢深静脉血栓 (Deep vein thrombosis of the lower extremities) – 皮肤紫色 (skin becomes purple)

- 心律失常患者可以表现以心悸为主，可伴有胸闷……(Patients with arrhythmia may mainly manifest with palpitations, optionally accompanied by chest tightness . . .)
  ⇒ 心律失常 (arrhythmia) – 心悸 (palpitations); 心律失常 (arrhythmia) – 胸闷 (chest tightness)

Status–Require–Information The Status entity requires the Information entity to confirm or exclude,[4] where the Information is supposed to be unknown. Accordingly, the context should not explicitly mention the result (i.e., confirming or excluding the Status entity), since the information is still unknown; otherwise, the relation should be annotated as the Information–Suggest–Status or Information–Exclude–Status type described below.

- 消化性溃疡可通过查胃镜助诊 (Peptic ulcer can be diagnosed with gastroscopy)
  ⇒ 消化性溃疡 (Peptic ulcer) – 胃镜 (gastroscopy)
- 颈椎骨折脱位：此类患者有明确外伤史，起病急，X 线摄片或 CT 检查可鉴别 (Cervical fracture and dislocation: Such patients have a clear history of trauma and have a rapid onset, which can be distinguished by X-ray or CT examination)
  ⇒ 颈椎骨折脱位 (Cervical fracture and dislocation) – X 线摄片 (X-ray); 颈椎骨折脱位 (Cervical fracture and dislocation) – CT 检查 (CT examination)
- 患者出现腹痛、乏力，查全腹部 CT 显示……(The patient developed abdominal pain and fatigue. The CT scan of the entire abdomen showed . . .)
  ⇒ 腹痛 (abdominal pain) – 全腹部 CT (CT scan of the entire abdomen); 乏力 (fatigue) – 全腹部 CT (CT scan of the entire abdomen)

Information–Suggest–Status The Information entity suggests the existence of the Status entity, where the Information could be either positive or negative.

- 患者左面部疼痛，于医院诊断为三叉神经痛 (The patient' s left facial pain was diagnosed as trigeminal neuralgia in the hospital)
  ⇒ 左面部疼痛 (left facial pain) – 三叉神经痛 (trigeminal neuralgia)
- 患者目前黄疸型肝炎诊断明确，考虑药物性肝损可能性大 (The patient' s current diagnosis of jaundice hepatitis is clear, and the possibility of drug-induced liver damage should be considered)
  ⇒ 黄疸型肝炎 (jaundice hepatitis) – 药物性肝损 (drug-induced liver damage)
- 腰椎 MRI 示：L4/5、L5/S1 椎间盘突出……(Lumbar MRI showed: L4/5, L5/S1 intervertebral disc herniation . . .)
  ⇒ 腰椎 MRI (Lumbar MRI) – L4/5、L5/S1 椎间盘突出 (L4/5, L5/S1 intervertebral disc herniation)

Information–Exclude–Status The Information entity excludes the possibility of the Status entity, where the Information could be either positive or negative.

- 消化性溃疡：常有反复上腹隐痛，伴反酸，烧心等症，可引起呕血黑便，该患者已作胃镜排除，故不考虑 (Peptic ulcer: often with repeated upper abdominal pain, accompanied by acid reflux, heartburn, etc., which can cause hematemesis and melena. According to the patient's gastroscopy result, it has been excluded and not considered)
  ⇒ 胃镜 (gastroscopy) – 消化性溃疡 (Peptic ulcer)
- 多发性肌炎或皮肌炎：混合性结缔组织病患者一般无肌痛症状，肌酶大多正常 (Polymyositis or dermatomyositis: patients with mixed connective tissue disease generally have no symptoms of myalgia, and muscle enzymes are mostly normal)
  ⇒ 肌痛 (myalgia) – 多发性肌炎 (Polymyositis); 肌痛 (myalgia) – 皮肌炎 (dermatomyositis)

3.2 Status vs. Intervention

Status–Require–Intervention The Status entity requires the Intervention entity to modify, such as a disease requiring treatment. The context should not explicitly mention the result of intervention; otherwise, the relation should be annotated as the Intervention–Modify–Status type described below.

- 患者右锁骨骨折伴右肩锁关节脱位，在我院住院行手术治疗 (The patient had a fracture of the right clavicle and dislocation of the right acromioclavicular joint, and was hospitalized in our hospital for

[4] Or the Status entity requires the Information entity to confirm its cause.

surgical treatment)
⇒ 右锁骨骨折 (fracture of the right clavicle) – 手术治疗 (surgical treatment); 右肩锁关节脱位 (dislocation of the right acromioclavicular joint) – 手术治疗 (surgical treatment)

- 既往患高血压病史 26 年余，平素服用 “马来酸左旋氨氯地平分散片、缬沙坦” 控制血压 (With a history of hypertension for more than 26 years, the patient usually takes “levoamlodipine maleate dispersible tablets and valsartan” to control blood pressure)
⇒ 高血压 (hypertension) – 马来酸左旋氨氯地平分散片 (levoamlodipine maleate dispersible tablets); 高血压 (hypertension) – 缬沙坦 (valsartan)
- 患者左髌骨骨折术后 1 年余，要求拆除内固定 (The patient had left patella fracture surgery more than 1 year ago, and asks to remove the internal fixation)
⇒ 左髌骨骨折术后 (had left patella fracture surgery) – 拆除内固定 (remove the internal fixation)

Intervention–Modify–Status The Intervention entity modifies the Status entity. According to the result, this relation type can be further divided into Intervention–Improve–Status and Intervention–Worsen–Status.[5]

- 患者被诊断为三叉神经痛，在我院住院治疗（三叉神经神经射频热凝术）后，疼痛好转 (The patient was diagnosed with trigeminal neuralgia. After being hospitalized in our hospital (radiofrequency thermocoagulation of the trigeminal nerve), the pain was alleviated)
⇒ 三叉神经神经射频热凝术 (radiofrequency thermocoagulation of the trigeminal nerve) – 三叉神经痛 (trigeminal neuralgia)
- 患者规律服用阿卡波糖片 50mg tid，自述血糖控制可 (The patient takes acarbose tablets 50mg tid regularly, and says that blood sugar is well controlled)
⇒ 阿卡波糖片 (acarbose tablets) – 血糖控制可 (blood sugar is well controlled)
- 予 “依托考昔片” 对症治疗后，疼痛仍存，较前稍缓解 (After symptomatic treatment with “Etocoxib Tablets”, the pain still persists but is slightly relieved than before)
⇒ 依托考昔片 (Etocoxib Tablets) – 疼痛 (pain)
- 该患者诊断明确，神经症状明显，保守治疗效果不佳 (The patient has a clear diagnosis, obvious neurological symptoms, so conservative treatment should be ineffective)
⇒ 保守治疗 (conservative treatment) – 神经症状 (neurological symptoms)

Intervention–Cause–Status The Intervention entity causes the Status entity, which is typically a complication.

- 服用倍他乐克/比索洛尔期间注意心率、血压，如有头晕等症状……(Pay attention to the heart rate and blood pressure while taking Betaloc/Bisoprolol. If there are symptoms like dizziness . . .)
⇒ 倍他乐克 (Betaloc) – 头晕 (dizziness); 比索洛尔 (Bisoprolol) – 头晕 (dizziness)
- 患者化疗后出现骨髓抑制，感染性发热，咳嗽咳痰明显 (The patient developed bone marrow suppression, infectious fever, and obvious cough and sputum after chemotherapy)
⇒ 化疗 (chemotherapy) – 骨髓抑制 (bone marrow suppression); 化疗 (chemotherapy) – 感染性发热 (infectious fever); 化疗 (chemotherapy) – 咳嗽 (cough); 化疗 (chemotherapy) – 咳痰 (sputum)

3.3 Intervention vs. Information

Intervention–Require–Information The Intervention entity requires the Information entity to confirm its feasibility, where the Information is supposed to be unknown. Accordingly, the context should not explicitly mention the confirming result, since the information is still unknown; otherwise, the relation should be annotated as the Information–Permit–Intervention or Information–Contraindicate–Intervention type described below.

- 择日进行冠脉造影 + 冠状动脉介入治疗 (Schedule for coronary angiography + coronary intervention)
⇒ 冠状动脉介入 (coronary intervention) – 冠脉造影 (coronary angiography)
- 损伤处查 CT 或 MRI，明确病情，排除手术禁忌 (Check CT or MRI for the injury to clarify the condition and exclude surgical contraindications)
⇒CT (CT) – 手术 (surgical); MRI (MRI) – 手术 (surgical)

[5]It is also annotated as Intervention–Worsen–Status if the intervention is ineffective.

Table 2: Medical relation types and applicable entity types

Relation type	Head entity types	Tail entity types
Status–Cause–Information	Disease, Symptom, Personal History	Disease, Symptom
Status–Require–Information	Disease, Symptom	Test Process
Information–Suggest–Status	Disease, Symptom, Test Process, Test Result	Disease, Symptom
Information–Exclude–Status	Disease, Symptom, Test Process, Test Result	Disease, Symptom
Status–Require–Intervention	Disease, Symptom	Treatment, Drug
Intervention–Modify–Status	Treatment, Drug	Disease, Symptom
Intervention–Cause–Status	Treatment, Drug	Disease, Symptom
Intervention–Require–Information	Treatment, Drug	Test Process
Information–Permit–Intervention	Disease, Symptom, Test Process, Test Result	Treatment, Drug
Information–Contra–Intervention	Disease, Symptom, Test Process, Test Result	Treatment, Drug

Notes: If an entity type is listed, the corresponding relations apply to all its sub types.

- 服用他汀类降脂药物期间，定期随访肝功能、血脂、肌酶 (During the period of taking statin lipid-lowering drugs, regular follow-up testing liver function, blood lipids and muscle enzymes)
  ⇒ 他汀类降脂药物 (statin lipid-lowering drugs) – 肝功能 (liver function); 他汀类降脂药物 (statin lipid-lowering drugs) – 血脂 (blood lipids); 他汀类降脂药物 (statin lipid-lowering drugs) – 肌酶 (muscle enzymes)

Information–Permit–Intervention The Information entity confirms the feasibility of the Intervention entity, where the Information could be either positive or negative.

- 现患者血常规、肝肾功能正常，今起可按吉西他滨针 1.6g D1，D8 方案化疗 (Now the patient' s blood routine, liver and kidney functions are normal. From now on, he can follow the 1.6g D1 and D8 regimen of gemcitabine for chemotherapy)
  ⇒ 血常规 (blood routine) – 化疗 (chemotherapy); 肝肾功能 (liver and kidney functions) – 化疗 (chemotherapy)
- 血常规、大生化、凝血功能、胸片、心电图等未见手术禁忌 (There are no contraindications to surgery according to blood routine, biochemistry, coagulation function, chest radiograph, electrocardiogram, etc)
  ⇒ 血常规 (blood routine) – 手术 (surgery); 大生化 (biochemistry) – 手术 (surgery); 凝血功能 (coagulation function) – 手术 (surgery); 胸片 (chest radiograph) – 手术 (surgery); 心电图 (electrocardiogram) – 手术 (surgery)
- 如无明显低血糖，必要时可使用瑞格列奈片 (If there is no obvious hypoglycemia, repaglinide tablets can be used if necessary)
  ⇒ 无明显低血糖 (no obvious hypoglycemia) – 瑞格列奈片 (repaglinide tablets)

Information–Contraindicate–Intervention The Information entity contraindicates the application of the Intervention entity, where the Information could be either positive or negative.

- CT 无法明确显示 L L5-S1 椎间盘平面，经与家属商议后放弃椎间盘穿刺射频治疗 (CT cannot clearly show the plane of the LL5-S1 intervertebral disc, and after discussing with the patient's family members, the radiofrequency treatment of intervertebral disc puncture is abandoned)
  ⇒CT (CT) – 椎间盘穿刺射频治疗 (radiofrequency treatment of intervertebral disc puncture)
- 家属内固定取出后有再骨折风险，患者及家属表示理解，暂不考虑手术 (There is a risk of re-fracture after the family member' s internal fixation is taken out. The patient and family members express their understanding and do not consider surgery for the time being)
  ⇒ 再骨折 (re-fracture) – 手术 (surgery)

3.4 Summary

Table 2 presents the ten medical relation types defined in our scheme, and their applicability to entity types.

Literalness is prior to external knowledge Entity relations are high-level semantics, so ideally, the confident annotation of a medical relation relies on both the correct text expression and the consistency with

external knowledge. However, as reported by the annotators, the real clinical texts contain a non-negligible quantity of typos, grammar mistakes and content errors, which may make it difficult to interpret the relations or cause conflicts between text meanings and external knowledge. As previously highlighted, the annotations should respect the text literalness more preferentially than the external knowledge. Hence, as long as the text clearly states a medical relation, the annotation should follow it; otherwise, the annotator can try to infer implicit relations from the text with the help of her medical knowledge.

Determining the entity roles first In our scheme, a Disease or Symptom entity may play a Status role or a Information role in a medical relation, depending on the contexts. Hence, given two entities, the annotator should first choose a role for each entity and then label their relation. In the case of two Disease (or Symptom) entities appearing in a relation, the annotator has to decide which is the relative essence and thus plays the Status role, and which is the relative appearance and thus plays the Information role.

Annotating relations regardless of attributes The relation annotation should be based on the original versions of entities, regardless of their attributes [He et al., 2017]. Note that some relation types are antonyms to each other, such as Information–Suggest–Status versus Information–Exclude–Status, Information–Permit–Intervention versus Information–Contra–Intervention. It would be confusing to annotate such relations on entities with a Negation attribute (see Subsection 4.1). For example, the text may state that a negated Information entity suggests a Status entity, which exactly means that the original Information entity excludes the Status entity. In this case, either Information–Suggest–Status or Information–Exclude–Status relation can be justified. As an agreement, we chose to not consider any attributes when annotating relations.

4 Medical Attribute

When annotating an entity in clinical text, we assume by default that the underlying disease (or entity of other types) is currently present on the patient, without explicit uncertainty, conditionality or occasionality, or its status being better or worse. However, the context may deviate from the default. In such case, we would annotate specific attributes on the entity to indicate such deviation.

Given tokens x and entities y^{Ent}, the objective output of attribute extraction is a set y^{Attr}, of which each element is a doublet of attribute type and entity:

$$y^{Attr} = \{(type_k^{Attr}, entity_k) \mid type_k^{Attr} \in S^{Attr}, entity_k \in y^{Ent}, k = 1, 2, \ldots, K\} \tag{3}$$

where S^{Attr} is the set of attribute types, K is the number of attributes of the entities.[6] Hence, the annotation of a medical attribute should specify its attribute type and associated entity.

We formulate ten attribute types: Negation, Family, Analysis, Uncertainty, Conditionality, Occasionality, Better, Worse, History, Future.

4.1 Negation

A Negation attribute indicates that the corresponding entity does not exist, or has not been used. An exceptional policy is that Test Result and Abnormal Test Result entities should not be attached with Negation attributes. Since Test Result is defined to be normal, Test Result and Abnormal Test Result are semantically the negated versions to each other. In other words, any Test Result entity with a Negation attribute is equivalent to a corresponding Abnormal Test Result entity, and *vice versa*. Hence, the annotation scheme formulates that any Test Result entity with a Negation attribute should be transformed to the equivalent Abnormal Test Result entity, and *vice versa*. This policy does not apply to other entity types like Self-Reported Abnormality or Test Process; they can be normally negated.

- ……否认心脏病、糖尿病史……(. . . denies heart disease or diabetes history . . .)
  ⇒ 心脏病 (heart disease), 糖尿病 (diabetes)
- ……无畏寒发热，无咳嗽咳痰……(. . . No fear of cold and fever, no cough and sputum . . .)
  ⇒ 畏寒 (fear of cold), 发热 (fever), 咳嗽 (cough), 咳痰 (sputum)
- 家属表示理解，暂不考虑手术 (Family members expressed their understanding and will not consider surgery for the time being)
  ⇒ 手术 (surgery)

[6] An entity may have no attribute, one attribute, or multiple attributes.

- 患者否认吸烟、饮酒史，否认吸毒史 (The patient denies the history of smoking and drinking, and denies the history of drug use)
  ⇒ 吸烟 (history of smoking), 饮酒史 (drinking), 吸毒史 (history of drug use)

4.2 Family

A Family attribute indicates that the corresponding entity exists on the patient's family members, rather than the patient.

- 大多数患者有偏头痛家族史 (Most patients have a family history of migraine)
  ⇒ 偏头痛 (migraine)

4.3 Analysis

An Analysis attribute indicates that the context is analysis regarding general medical theories and experiences, rather than specified to the current patient. This attribute typically appears in clinical analysis, such as prognosis and differential diagnosis. Note that the subject disease of a differential diagnosis should be attached with an Uncertainty attribute rather than an Analysis attribute, because it is regarding the patient (but not confirmed); while the entities in the description text should be attached with Analysis attributes.

- 鉴别诊断：1. 食管穿孔：可有胸前区、肩胛间区及剑突下疼痛……(Differential diagnosis: 1. Esophageal perforation: there may be pain in the anterior chest area, interscapular area and under the xiphoid process . . .)
  ⇒ 胸前区、肩胛间区及剑突下疼痛 (pain in the anterior chest area, interscapular area and under the xiphoid process)
- 该疾病见于老年人，疼痛逐渐加重，影像学检查可见骨破坏且累及椎弓根……(The disease typically appears on the elderly, the pain gradually worsens, imaging examination would show bone destruction involving the pedicle . . .)
  ⇒ 疼痛 (pain), 影像学检查 (imaging examination), 骨破坏且累及椎弓根 (bone destruction involving the pedicle)
- 术后注意预防感染及深静脉血栓 (Pay attention to preventing infection and deep vein thrombosis after operation)
  ⇒ 感染 (infection), 深静脉血栓 (deep vein thrombosis)
- 若出现消化道穿孔，可考虑急诊手术治疗 (If perforation of the digestive tract occurs, consider emergency surgery)
  ⇒ 消化道穿孔 (perforation of the digestive tract), 急诊手术 (emergency surgery)
- 如情况允许，择期行“内固定拆除术”，并告知手术风险 (If the situation permits, schedule for "internal fixation removal" and inform the risk of the operation)
  ⇒ 内固定拆除术 (internal fixation removal)

4.4 Uncertainty

An Uncertainty attribute indicates that the context expresses significant uncertainty regarding the corresponding entity. As aforementioned, a typical example is that the subject disease of a differential diagnosis should be attached with an Uncertainty attribute.

- 鉴别诊断：1. 食管穿孔：可有胸前区、肩胛间区及剑突下疼痛……(Differential diagnosis: 1. Esophageal perforation: there may be pain in the anterior chest area, interscapular area and under the xiphoid process . . .)
  ⇒ 食管穿孔 (Esophageal perforation)
- 目前诊断考虑：肝功能不全：病毒性肝炎？药物性肝损？(Current diagnosis considerations: liver insufficiency: viral hepatitis? Drug-induced liver damage?)
  ⇒ 肝功能不全 (liver insufficiency), 病毒性肝炎 (viral hepatitis), 药物性肝损 (Drug-induced liver damage)
- 左侧附件区囊状低密度，不除外感染可能，必要时复查 (The cyst is of low density in the left appendage area, which does not exclude the possibility of infection, retest if necessary)
  ⇒ 感染 (infection)

- 恶性肿瘤可能性大 (Malignant tumors are probable)
  ⇒ 恶性肿瘤 (Malignant tumors)

4.5 Conditionality

A Conditionality attribute indicates that the underlying symptom only occurs in specific conditions, rather than always exists.

- ……活动后略感气促，无咳嗽咳痰，偶有畏寒，无发热寒战……(. . . slight short of breath after the activity, no cough or sputum, occasional chills, no fever and chills . . .)
  ⇒ 气促 (short of breath)
- 进食硬食物后有胸骨后胀痛，偶有反酸、嗳气 (After eating hard food, there is pain in the back of the sternum, acid reflux and belching)
  ⇒ 胸骨后胀痛 (pain in the back of the sternum)
- 消化道症状不如病毒性肝炎明显，停药后肝功能可明显好转或完全恢复 (Digestive tract symptoms are less obvious than viral hepatitis, the liver function can be significantly improved or completely restored after stopping the drug use)
  ⇒ 消化道症状 (Digestive tract symptoms)
- 患者右足跟轻微疼痛存在，跑步或者起跳后疼痛明显加重 (There is slight pain in the right heel of the patient, and the pain significantly worsens after running or jumping)
  ⇒ 右足跟轻微疼痛 (slight pain in the right heel)

4.6 Occasionality

An Occasionality attribute indicates that the underlying symptom occurs occasionally, rather than always exists.

- ……活动后略感气促，无咳嗽咳痰，偶有畏寒，无发热寒战……(. . . slight short of breath after the activity, no cough or sputum, occasional chills, no fever and chills . . .)
  ⇒ 畏寒 (chills)
- 进食硬食物后有胸骨后胀痛，偶有反酸、嗳气 (After eating hard food, there is pain in the back of the sternum, acid reflux and belching)
  ⇒ 反酸 (acid reflux), 嗳气 (belching)
- 患者诉清晨偶有胸闷气短，可自行缓解 (The patient complains of occasional chest tightness and breath shortness in the morning, which can be relieved by self)
  ⇒ 胸闷 (chest tightness), 气短 (breath shortness)

4.7 Better and Worse

A Better (Worse) attribute indicates that the context explicitly expresses that the underlying disease or symptom is getting better (worse).

- 患者被诊断为三叉神经痛，在我院住院治疗（三叉神经神经射频热凝术）后，疼痛好转 (The patient was diagnosed with trigeminal neuralgia. After being hospitalized in our hospital (radiofrequency thermocoagulation of the trigeminal nerve), the pain was alleviated)
  ⇒ 三叉神经痛 (trigeminal neuralgia)
- 手术前放疗可使癌块缩小，提高切除率和存活率 (Radiotherapy before surgery can shrink the tumor mass, improve the resection rate and survival rate)
  ⇒ 癌块 (tumor mass)
- 予“依托考昔片”对症治疗后，疼痛仍存，较前稍缓解 (After symptomatic treatment with “Etocoxib Tablets”, the pain still persists but is slightly relieved than before)
  ⇒ 疼痛 (pain)
- 两肺多发炎症改变，对照前片增多 (Multiple inflammatory changes in both lungs, increased in contrast to the previous images)
  ⇒ 两肺多发炎症改变 (Multiple inflammatory changes in both lungs)

4.8 History and Future

A History (Future) attribute indicates that the corresponding entity exists in the past (can be anticipated in the future), rather than in the current visit.

- 患者曾到医院就医，查胸部 CT 示胸腔积液增多，考虑结核性胸膜炎可能 (The patient went to the hospital for medical treatment, and the CT scan of the chest showed increased pleural effusion. Consider the possibility of tuberculosis)
  ⇒ 胸部 CT (CT scan of the chest), 胸腔积液增多 (increased pleural effusion), 结核性胸膜炎 (tuberculosis)
- 患者 1 周前在家中无明显诱因下出现胸闷气促 (The patient had chest tightness and shortness of breath at home without obvious cause 1 week ago)
  ⇒ 胸闷 (chest tightness), 气促 (shortness of breath)
- 患者需定期复查肝肾功能、血常规、血沉 (Patients need to regularly retest liver and kidney functions, blood routine, erythrocyte sedimentation rate)
  ⇒ 肝肾功能 (liver and kidney functions), 血常规 (blood routine), 血沉 (erythrocyte sedimentation rate)
- 疾病进展可出现“消化道穿孔”等并发症 (Complications such as "digestive tract perforation" may occur during the disease)
  ⇒ 消化道穿孔 (digestive tract perforation)

4.9 Summary

Table 3 presents the ten medical attribute types and their applicability to entity types.

Annotating attributes according to explicit literalness Following the principle of respecting literalness, an attribute should be annotated only when the local context has explicitly and significantly expressed the corresponding status, which is, by definition, deviated from the default. For example, any diagnoses should be assumed to be uncertain before sufficient medical tests, while annotating all these Disease entities as Uncertain would yield many trivial and redundant annotations. Besides, they are inferred from the global context, and provide limited incremental information regarding the local text. Hence, we chose to annotate attributes based on the local context.

Table 3: Medical attribute types and applicable entity types

	Negation	Family	Analysis	Uncertainty	Conditionality	Occasionality	Better & Worse	History & Future
Disease	✓	✓	✓	✓			✓	✓
Symptom								
– Self-Reported Abnormality	✓		✓		✓	✓	✓	✓
– Abnormal Test Result			✓		✓	✓	✓	✓
Test								
– Test Process	✓		✓					✓
– Test Result			✓					
Treatment	✓		✓					✓
Drug	✓		✓					✓
Body			✓					
Personal History	✓		✓					
Equipment								
Department								

Notes: If an entity type is listed, the corresponding attributes apply to all its sub types.

References

Leonardo Campillos, Louise Deléger, Cyril Grouin, Thierry Hamon, Anne-Laure Ligozat, and Aurélie Névéol. A French clinical corpus with comprehensive semantic annotations: development of the Medical Entity and Relation LIMSI annOtated Text corpus (MERLOT). *Language Resources and Evaluation*, 52(2):571–601, 2018.

Tongfeng Guan, Hongying Zan, Xiabing Zhou, Hongfei Xu, and Kunli Zhang. CMeIE: Construction and evaluation of Chinese medical information extraction dataset. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 270–282. Springer, 2020.

Bin He, Bin Dong, Yi Guan, Jinfeng Yang, Zhipeng Jiang, Qiubin Yu, Jianyi Cheng, and Chunyan Qu. Building a comprehensive syntactic and semantic corpus of Chinese clinical texts. *Journal of Biomedical Informatics*, 69:203–217, 2017.

Stéphane Meystre and Peter J Haug. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *Journal of Biomedical Informatics*, 39(6):589–599, 2006.

Philip Ogren, Guergana Savova, and Christopher Chute. Constructing evaluation corpora for automated clinical named entity recognition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 2008.

Angus Roberts, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42 (5):950–966, 2009.

Özlem Uzuner, Imre Solti, and Eithon Cadag. Extracting medication information from clinical text. *Journal of the American Medical Informatics Association*, 17(5):514–518, 2010.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5): 552–556, 2011.

Enwei Zhu, Qilin Sheng, Huanwan Yang, and Jinpeng Li. A unified framework of medical information annotation and extraction for Chinese clinical text. *Working Paper*, 2021.