Ubiquitous Computing Laboratory
Kyung Hee University, Korea

Thesis Presentation

Automatic Evidence Acquisition and Appraisal to support Evidence-based Decision Making

Thursday 3 November, 2016

Mr. Muhammad Afzal

Department of Computer Science and Engineering

Kyung Hee University, South Korea

Email: muhammad.afzal@oslab.khu.ac.kr

Advisor: Prof. Sungyoung Lee, PhD

Presentation Agenda

- Introduction
- Motivation and Objectives
- Problem Statement
- Proposed Methods
  - Solution 1 : Evidence Acquisition
  - Solution 2: Evidence Appraisal
- Experimental Results and Evaluations
- Uniqueness and Contributions
- Conclusion and Future Research

Introduction

- With the information explosion, the retrieval of the best clinical evidence from large and general purpose databases such as MEDLINE is difficult [Nancy et al 2005].

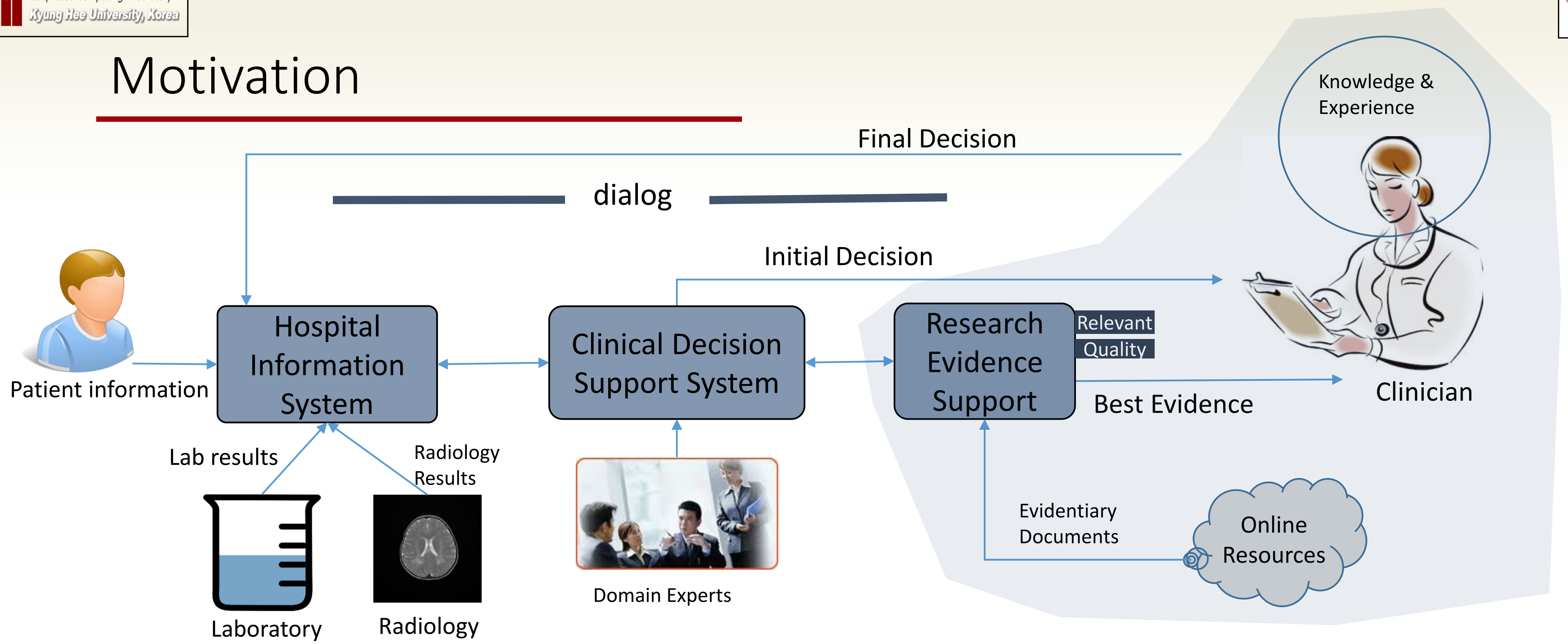
- Particularly in Evidence-based Medicine (EBM), the busy clinicians face numerous challenges to acquire best clinical evidence for quality care [Sackett, David L., et al 1996, Leung GM, 2001].

Decision Making Basic Elements

Clinical Expertise | Patients' Preferences | Research Evidence ⇒ The clinical evidence found in online available evidentiary documents.

"EBM, three legged stool"

Motivation



- Today number of MEDLINE Indexed articles
  - **21,508,439** (21 million+)
  - An internist require at least 20 scientific papers every day to keep up-to-date with this overwhelming number of yearly citations.

- Getting best available evidence is promising
  - Because, it will improve the confidence level of clinicians on clinical decisions
  - If made automatic, it will reduce unnecessary burden over clinicians/researchers

Problem Statement

In evidence-based medicine (EBM), without a well formulated question and an automated quality assessments, it is time consuming to identify a relevant and quality evidence [GRADEWG2004, Sarker2015, Boudin2010].

Goal

To minimize human efforts getting best research evidence for better clinical decision making.

Objectives

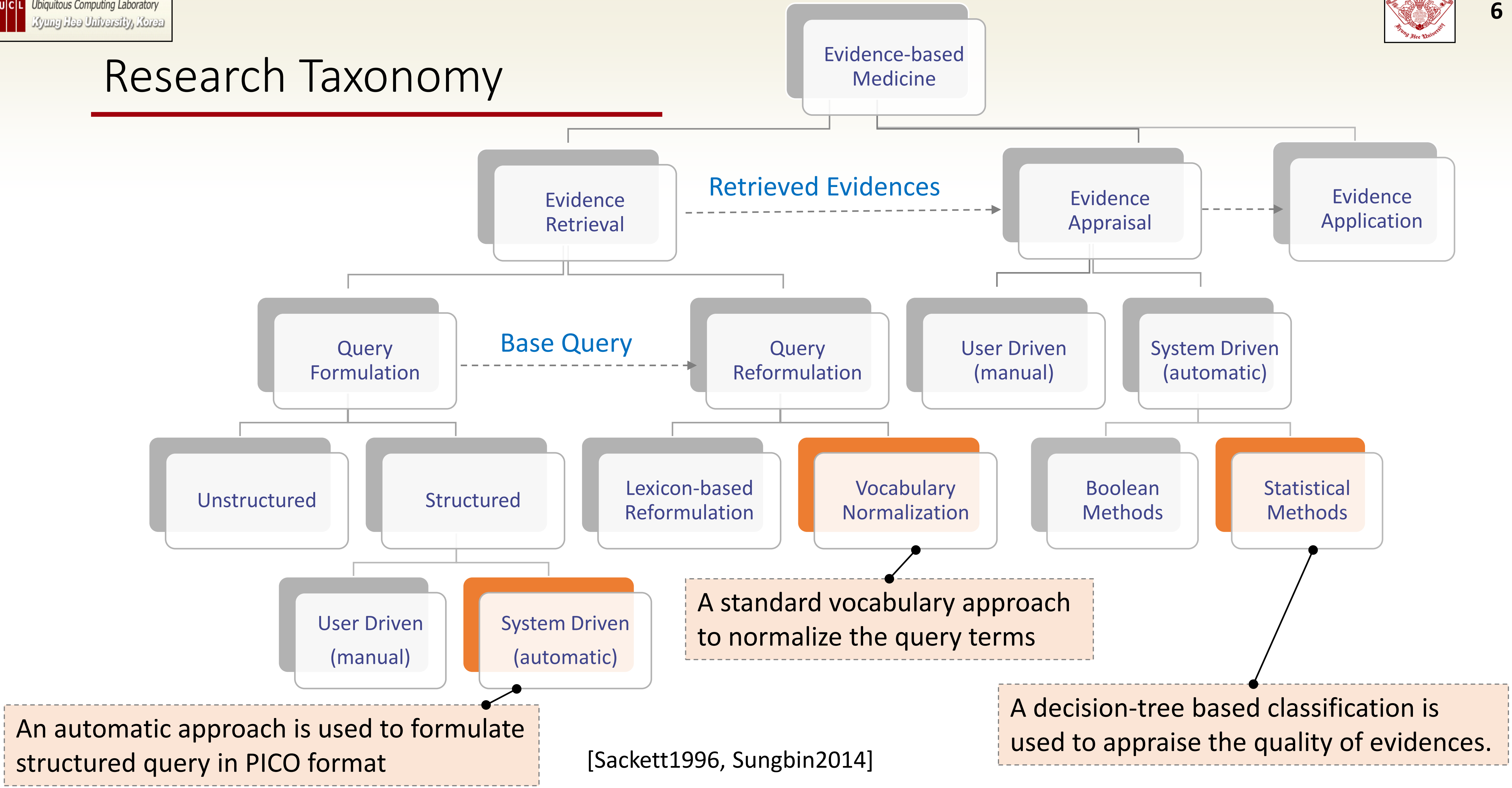
- To develop and evaluate methods/models for finding relevant evidentiary documents.

  Challenges: Retrieving task oriented relevant document with a higher precision

- To develop and evaluate methods/models for recognizing quality evidences.

  Challenges: Recognizing quality and contextually fit evidences with a higher accuracy

Research Taxonomy



[Sackett1996, Sungbin2014]

QF: Query Formulation
AQRF: Automatic Query Reformulation
SbQR: Statistical-based Quality Recognition
ERG: Evidence Ranking/Grading
CEG: Contextual Evidence Grading

Related Work (Individual Method-wise)

	Approach	QF	AQRF	SbQR	ERG	CEG
Pre-retrieval	Clinical Query [Wilczynski2005]	Yes (manual)	No	No	Yes (Ranking)	Yes(Partially)
	InfoButton [DelFiol2012]	Yes (semi-Auto)	No	No	No	Yes (manual)
	CDAPubMed[Perez2012]	Yes (semi-Auto)	Yes	No	No	No
	askMedline [Fontelo2005]	Yes (manual)	Yes	No	No	No
Post-retrieval	*Towards Automatic Recognition [Kilicoglu2009}*	No	No	Yes	No	No
	Evidence Quality Prediction [Sarker2015]	No	No	Yes	Yes (Grading)	No
	Proposed Approach	Yes (auto)	Yes	Yes	Yes (Grading)	Yes (auto)

Limitations

- Query building approaches are manual or semi-automatic
- Reformulation process consider terminological variants
- Dataset limitations and manual features engineering for quality evaluation statistically.
- Evidence grading without considering the user context
- Non-textual data consideration for quality evaluations
- Rule mining from the evidences

Limitation, Objectives and Proposed Solutions

Limitations

- Information driven, non-structured and manual construction of query
- Lack of target task awareness

- Boolean methods for quality assessments
- Lack of user-aware evidence grading

Objectives

Finding relevant evidentiary documents.

Recognize the quality of contents and contextual fitness

Solutions

Sol 1 — Automatic Evidence Acquisition

- A: PICO Compliant Question Construction
- B: Clinical Task Aware Query Formulation

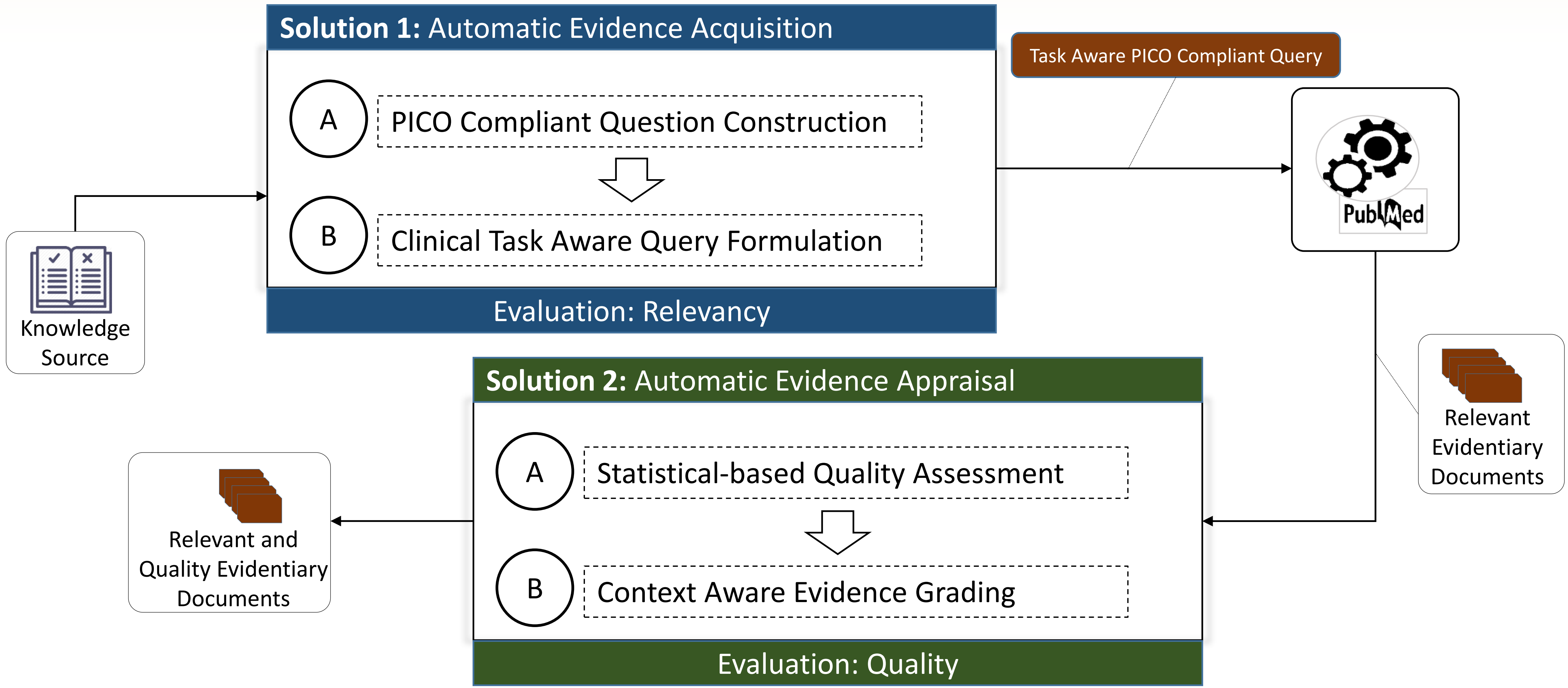
Sol 2 — Automatic Evidence Appraisal

- A: Statistical-based quality assessment
- B: User Aware contextual evidence grading

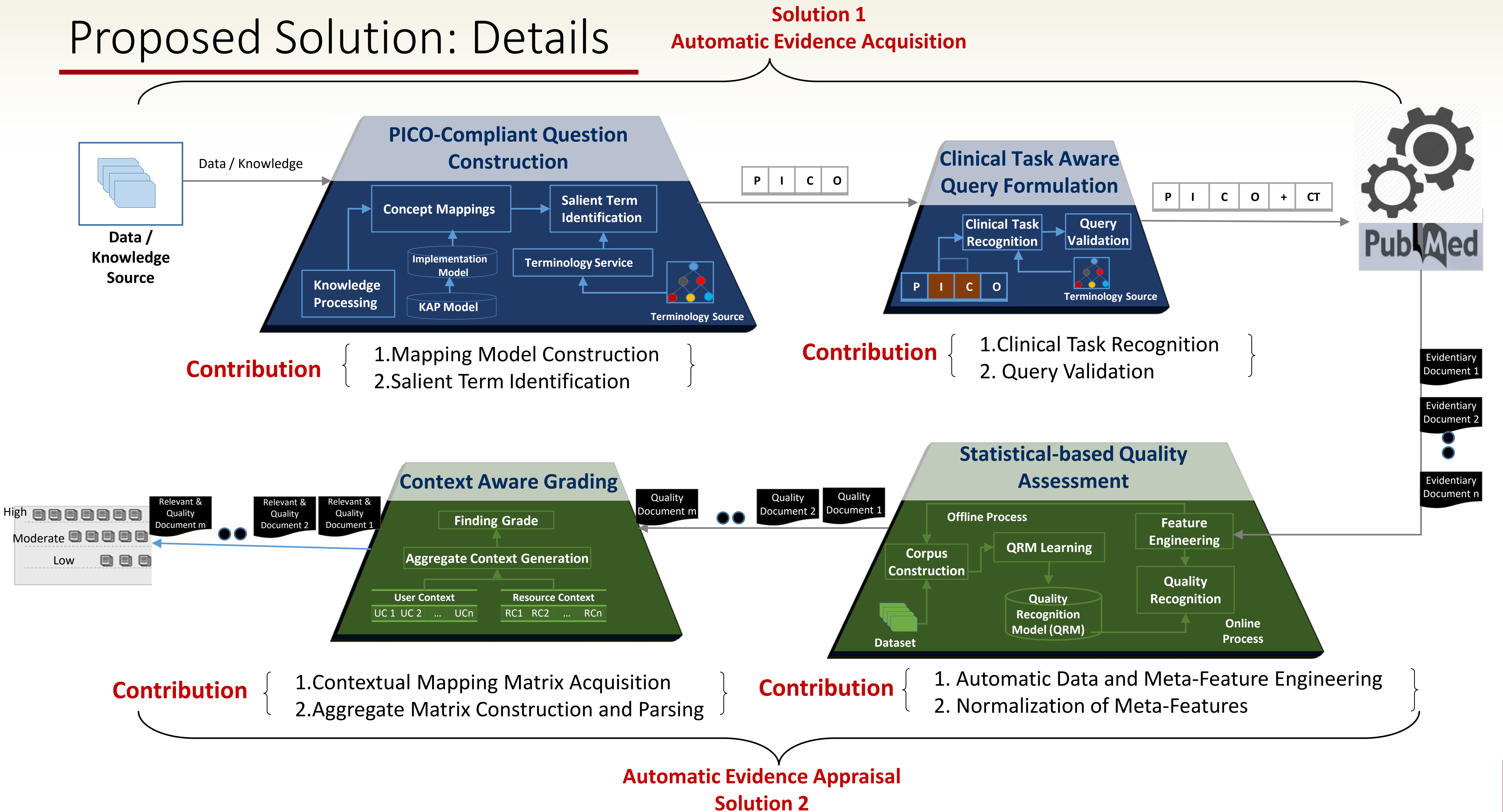
Goal

To minimize human efforts getting best research evidence for better clinical decision making.

Proposed Solution: Abstract Idea

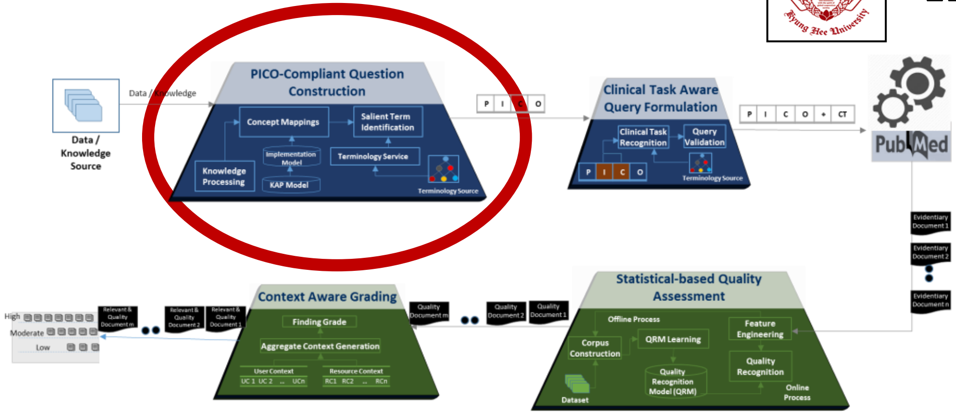


Proposed Solution: Details



Sol 1-A

PICO Compliant Question Construction (1/8)



Existing Approaches

Limitations

1. Unstructured →
   - Hard for human to understand the semantics of the query
   - Poor results are reported in EBM
2. Structured → Information Driven →
   - Manual
     - Time consuming
   - Automatic
     - Error prone

Proposed Approach

Knowledge Driven

Methods

- PICO compliant question construction
  - KAP Mapping Model
  - Salient Term Identification

Advantages

- Time efficient for the busy clinicians
- Better consistency irrespective of user expertise level.

What is PICO?

P	I	C	O
This include the primary problem, disease, or co-existing conditions.	This include intervention, prognostic factor, or exposure such as diagnostic test order, treatment plans.	This is an optional part of PICO which mainly include the alternative to intervention.	This include the goal to accomplish such as improving health of a patient, survivorship of a cancer patient etc.

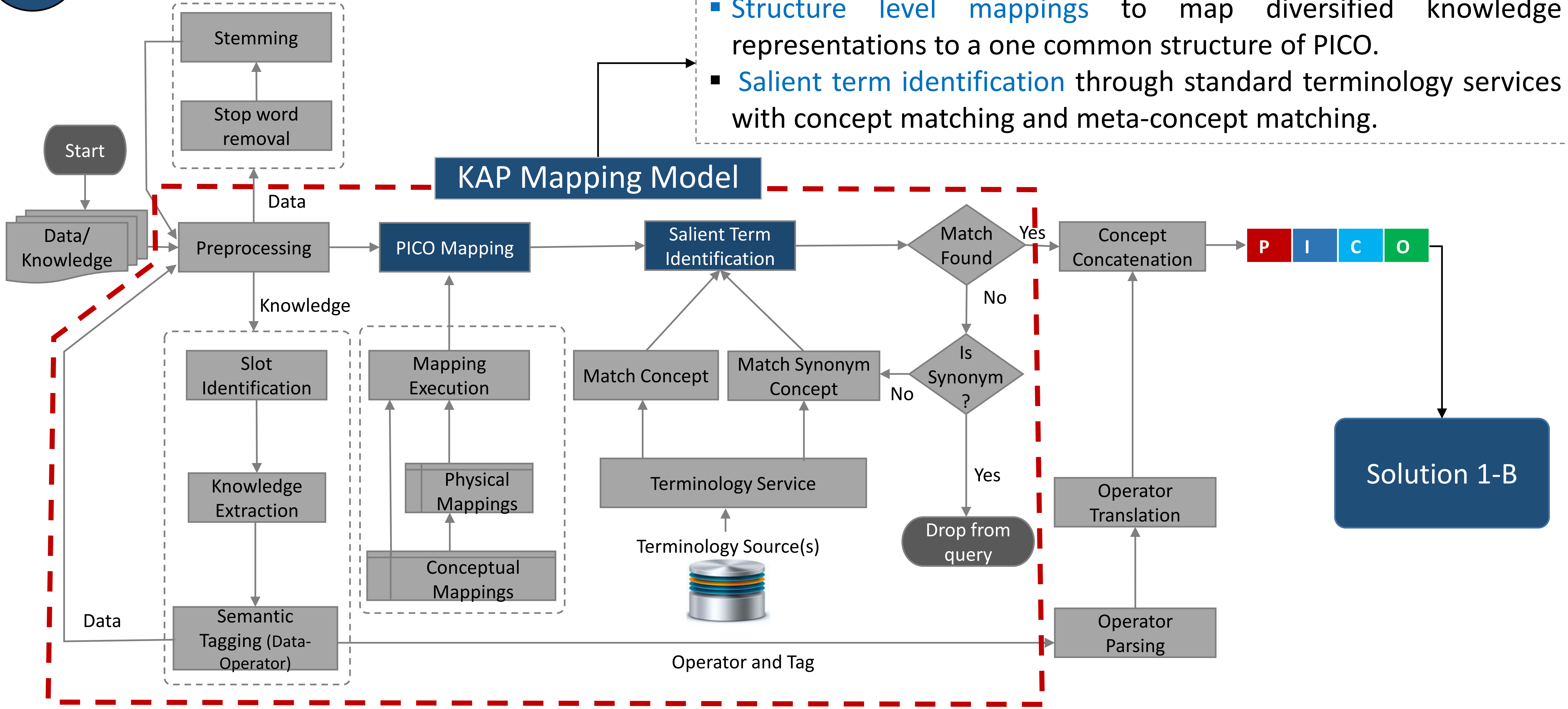
Why PICO?

[PICO2001, PICO2002]	[Hersh2008]	[Schardt2007]
Because, PICO facilitate the well-built search strategy based on four parts: (P), (I), (C), and (O), which are well matched with EBM Facets.	The PICO structure is commonly used in clinical studies.	Using a well-formulated question of PICO structure facilitates searching for a precise answer within a large medical citation database.

Sol 1-A

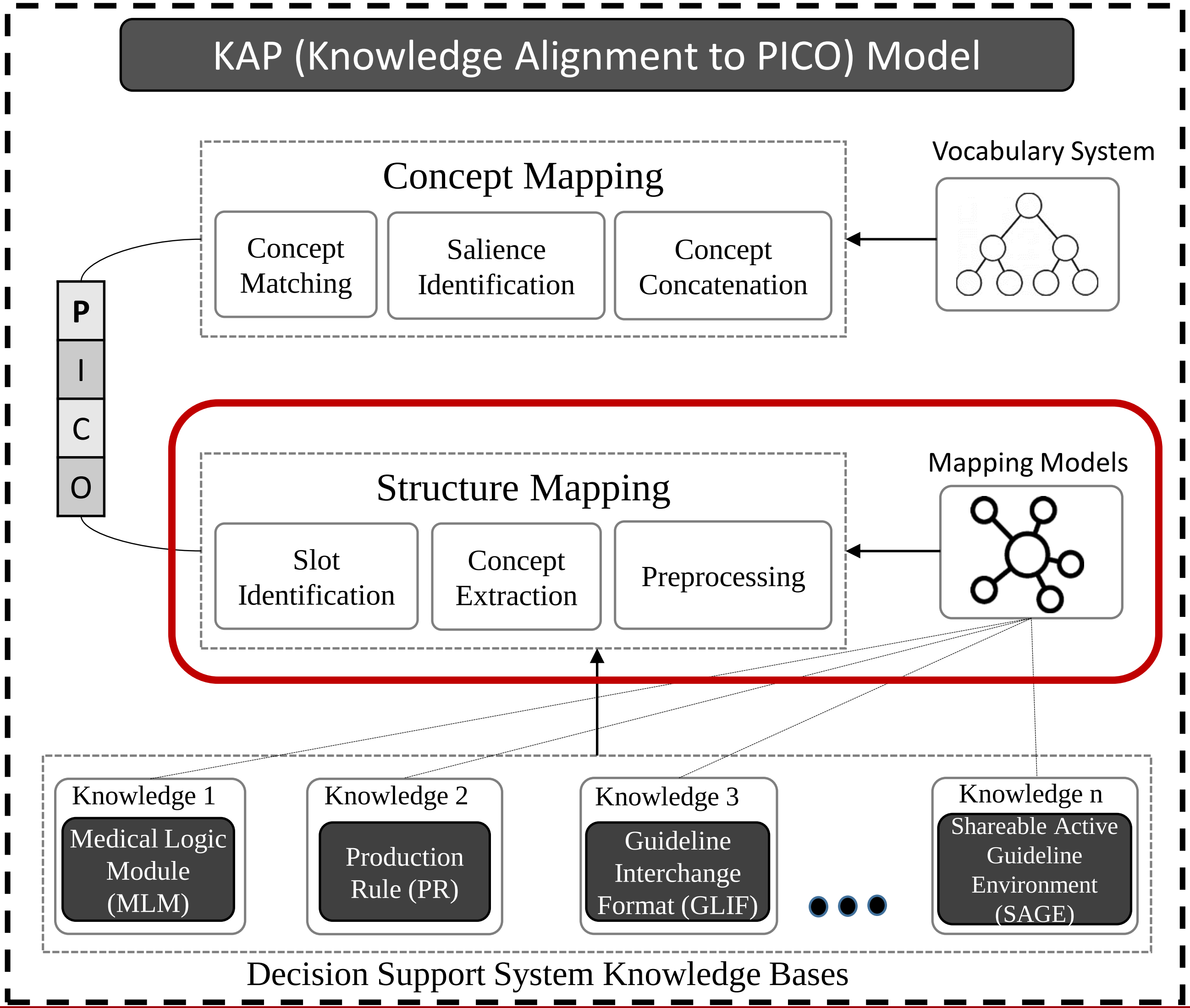
PICO Compliant Question Construction (2/8)

- Structure level mappings to map diversified knowledge representations to a one common structure of PICO.
- Salient term identification through standard terminology services with concept matching and meta-concept matching.



Sol 1-A PICO Compliant Question Construction (3/8)

Structure Mapping



$D \rightarrow P$
$A \rightarrow I$
$E \rightarrow C\ (Optional)$
$P \rightarrow O\ (Optional)$

Where;
D = *Rule Data* (*Conditoins*) concepts
A = *Rule Action* concepts
E = *Event* concepts
P = *Purpose* concepts

$$PICO = D \wedge A \wedge E \wedge P$$

$$PICO = \cap_{i=0}^{n} DC_i \wedge \cap_{i=0}^{n} AC_i \wedge \cap_{i=0}^{n} EC_i \wedge \cap_{i=0}^{n} PC_i \qquad (1)$$

$KB = R_1 \cup R_2 \cup \cdots \cup R_n$
$FRule \subseteq KB$
$\because FRule\ is\ the\ set\ of\ fired\ rules$

$$FRule = \begin{cases} R_1 \longrightarrow T_1 = \{t_1, t_2, \ldots, t_n\} \\ R_2 \longrightarrow T_2 = \{t_1, t_2, \ldots, t_n\} \\ \bullet\bullet\bullet \\ R_n \longrightarrow T_n = \{t_1, t_2, \ldots, t_n\} \end{cases}$$

$$QTerm = T_1 \cup T_2 \cup \cdots \cup T_n$$

$$QTerm = executedDecision\ Path$$

$$executedDecision = c \rightarrow d$$

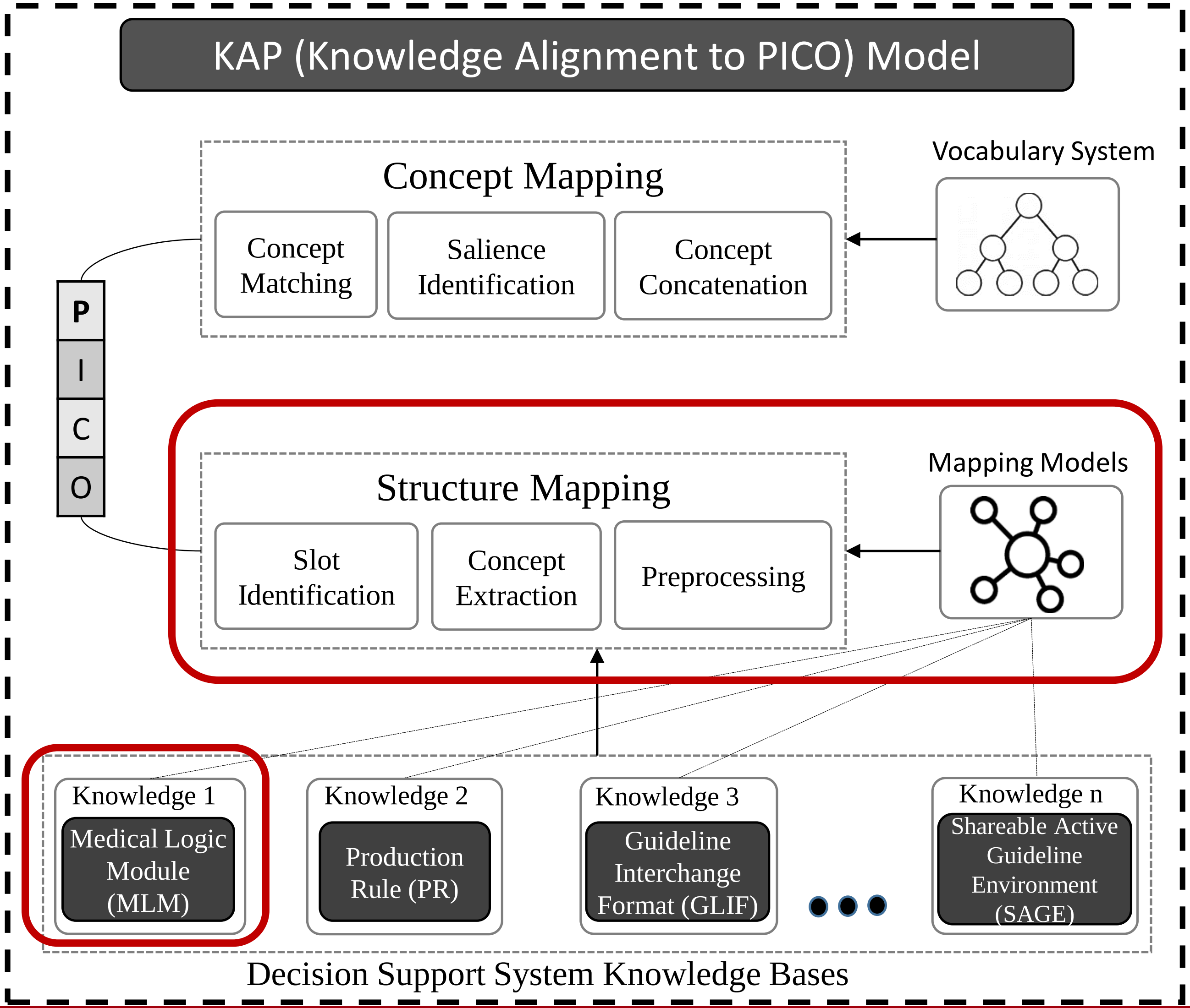
$$executedDecisionPath := \left\{ \begin{array}{l} p : decisoinPath;\ \exists r_1, r_2 \epsilon\ er \mid \\ (dom\ r_1 \cup r_2) \mapsto ran\ r_2 \Rightarrow \\ ran\ r_1 = \emptyset \wedge dom\ r_2 \subset (dom\ r_1 \cup dom\ r_2) \cdot \\ dom\ p = (dom\ r_1 \cup dom\ r_2) \wedge (ran\ p = ran\ r_2) \end{array} \right\}$$

(2)

Sol 1-A

PICO Compliant Question Construction (4/8)

Structure Mapping



Control Structure Parsing

- Resolving the semantics of different control structure used in logic of a rule (if-then, case, looping etc.)

MLM Control structure parsing rules

Example	Logic	Explanation
A	IF (C = "v1") THEN	**Condition sentence:** C = "v1"
	D = "d1"	
	Output: "d1 is recommended"	**Decision sentence:** D = "d1"
	END IF	
B	IF (C = "v1") THEN	For CDSS output "d1 is recommended":
	D = "d1"	**Condition sentence:** C = "v1"
	Output: "d1 is recommended"	**Decision sentence:** D = "d1"
	ELSE	For CDSS output "d2 is recommended":
	D = "d2"	**Condition sentence:** C != "v1"
	Output = "d2 is recommended"	Where "!" represents the negation (not).
	END IF	**Decision sentence:** D: d2
C	IF (C = "vl") THEN	For CDSS output "d1 is recommended":
	D = "d1"	**Condition sentence:** C = "v1"
	Output: "d1 is recommended"	**Decision sentence:** D = "d1"
	ELSEIF (C in ("v2", "v3")) THEN	For CDSS output "d2 is recommended":
	D = "d2"	**Condition sentence:** C in ("v2", "v3")
	Output: "d2 is recommended"	**Decision sentence:** D = "d2"
	ELSEIF (C = "v3") THEN	For CDSS Output "d3 is recommended"
	D = "d3"	**Condition sentence:** C = "v3"
	Output = "d3 is recommended"	**Decision sentence:** D = "d3"
	ELSE	For CDSS output "d4 is recommended"
	D = "d4"	**Condition sentence:** C != "v3"
	Output = "d4 is recommended"	**Decision sentence:** D = "d4"

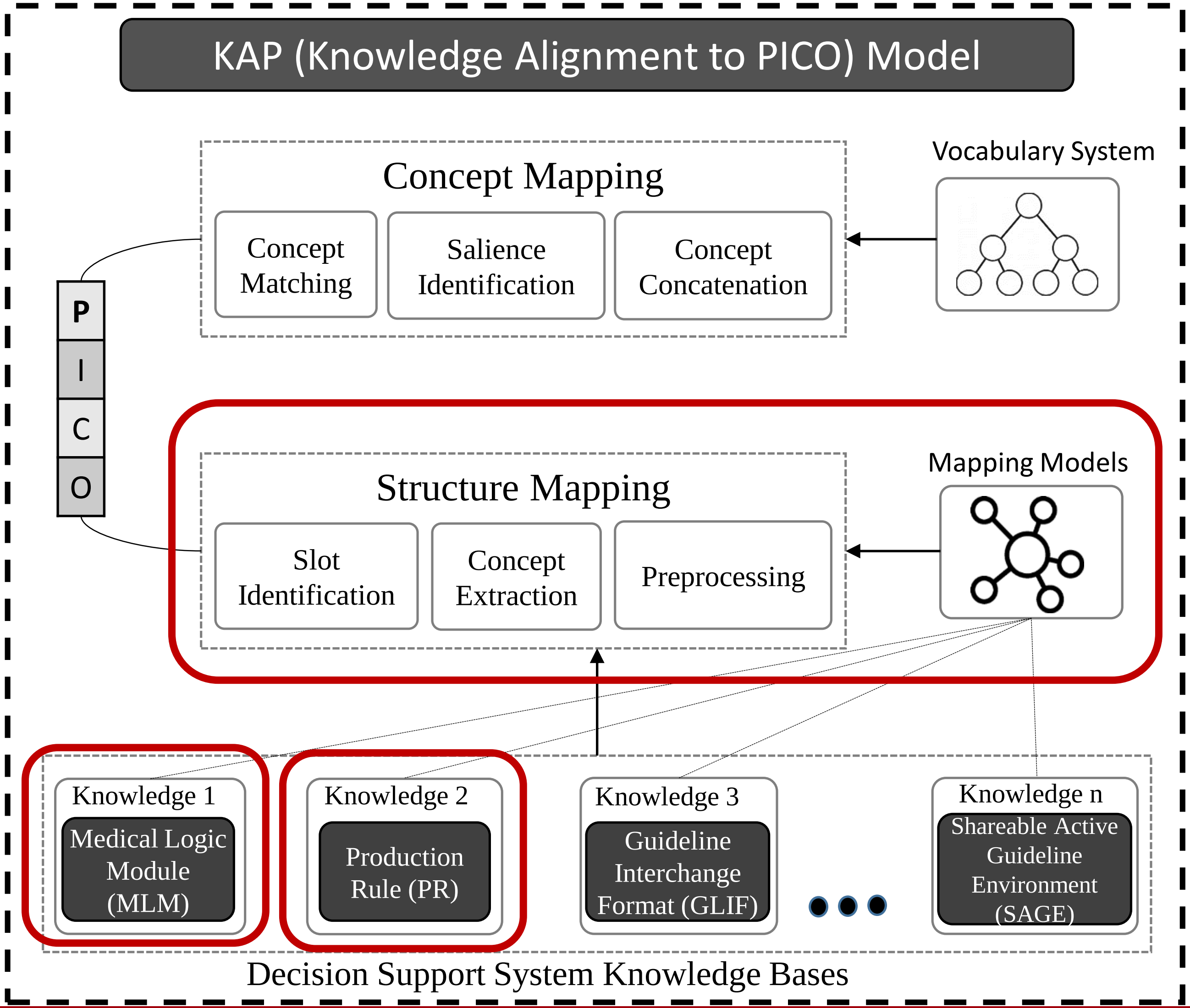
⋮

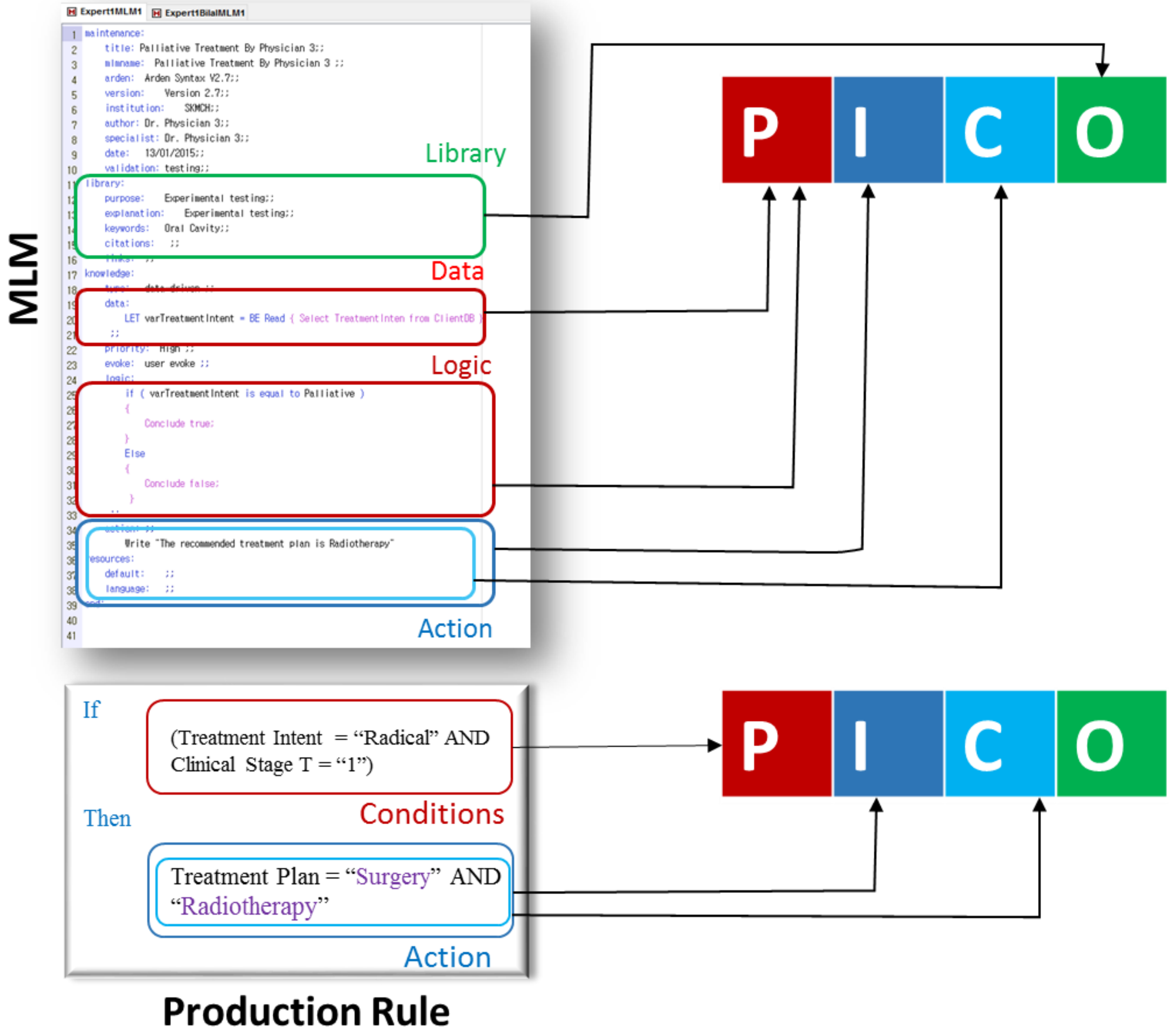
[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

Sol 1-A

PICO Compliant Question Construction (5/8)

Structure Mapping

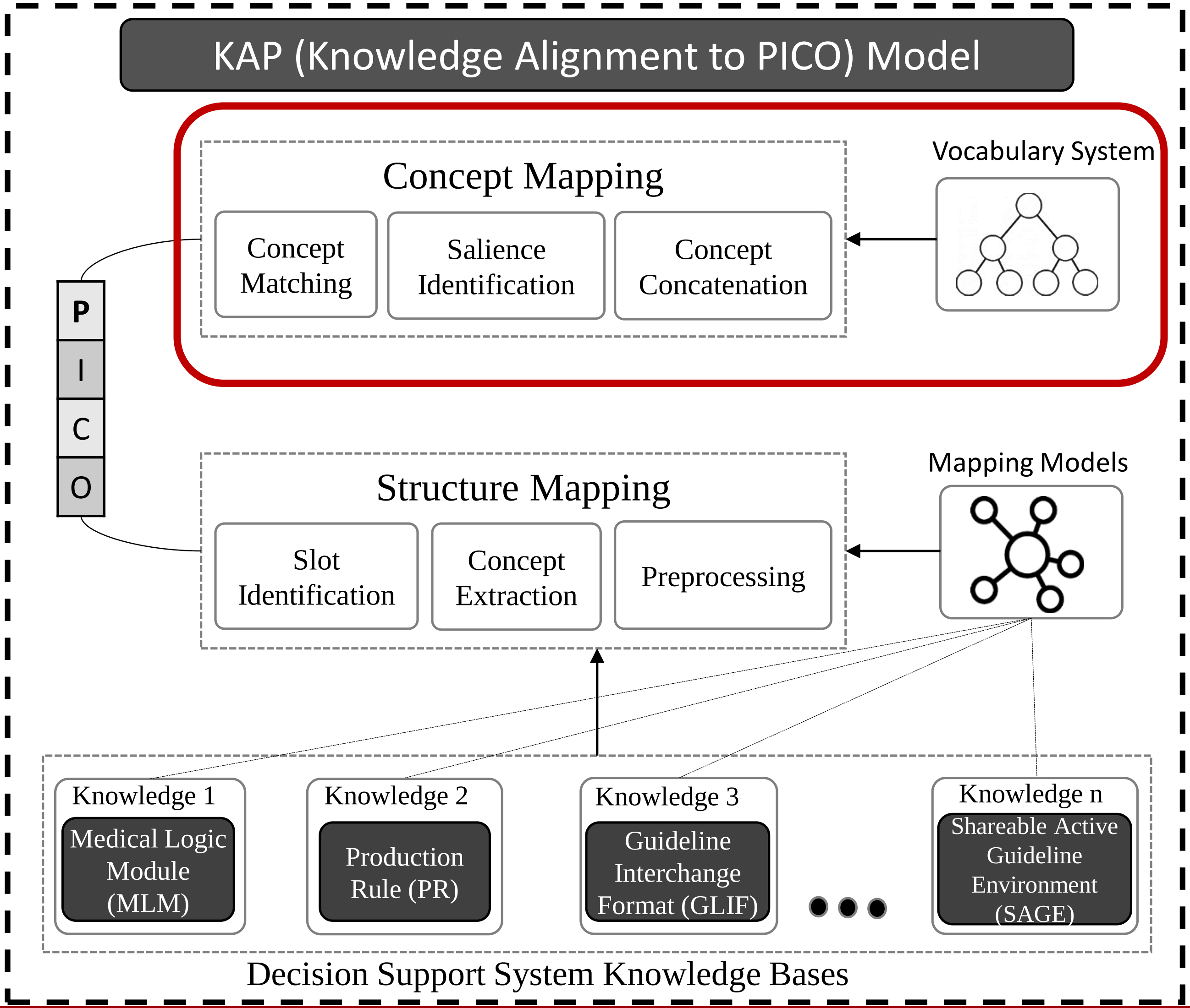




[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

Sol 1-A

PICO Compliant Question Construction (6/8)



Concept Mapping

***Existing Approach**: Consider every concept used in the query is important.*

***Proposed Approach**: Only a subset of the concepts is important.*

$$STerm \subseteq QTerm = D \cup A$$
$$STerm = STI(QTerm) \quad (3)$$

$Where, STI$: ***Salient Term Identification*** *algorithm*

String Matching

$$f(x) = C_i \bigcap C_j$$
$$and\ \forall C_i\ \varepsilon\ \mathrm{QTerm}, \forall C_j\ \varepsilon\ O \quad (3.1)$$

Synonym Matching

$$g(x) = C_i.labels \bigcap C_j.labels$$
$$and\ \forall C_i\ \varepsilon\ \mathrm{QTerm}, \forall C_i\ \varepsilon\ O \quad (3.2)$$

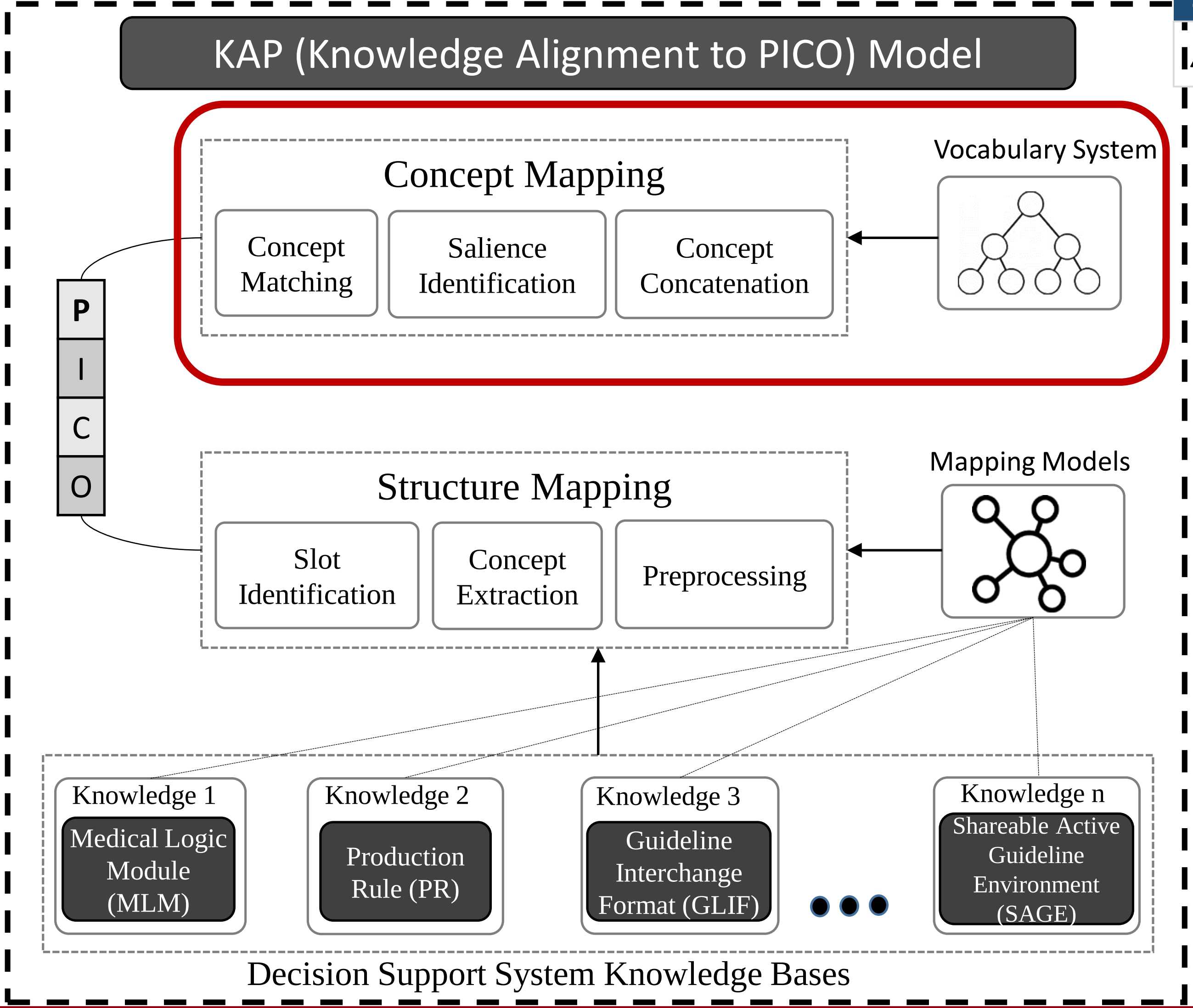
Where,
$\bigcap$ *is the set of string based matching techniques*
O is the ontology (e.g. SNOMED CT, UMLS)

[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

Sol 1-A

PICO Compliant Question Construction (7/8)

Concept Mapping



Salient Term Identification (STI)

***Approach A**: Consider every concept used in the query is important.*

***Approach B**: Only a subset of the concepts is important.*

Algorithm 1. Salient Terms Identification (STI)

Begin

inputs: $CTerm - \{t_1, t_2, \ldots, t_n\}$; //the list of m (*condition terms) extracted from rules*
output: $STerm - \{t_1, t_2, \ldots, t_m\}$; // the list of *m* (problem terms), where m ≤ n

STS; /* Where STS is the Terminology Service of SNOMED CT
for i = 0 to n-1
  if** (STS.Concepts.exist(i)) **then
    $parent \leftarrow getParent(i);$
    $\boldsymbol{if}\ (parent = \text{"}clinical\ finding\text{"})$
      PTerm.add(i);
    Endif
  Elseif** (STS.Synonyms.exist(i)) **then
    PTerm.add(i);
  Endif
Endfor
*return **STerm;***
End

Standard terminology service (STS).

UMLS → MetaMap Service and SemMap Service
SNOMED CT → The IHTSDO SNOMED CT Service

[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

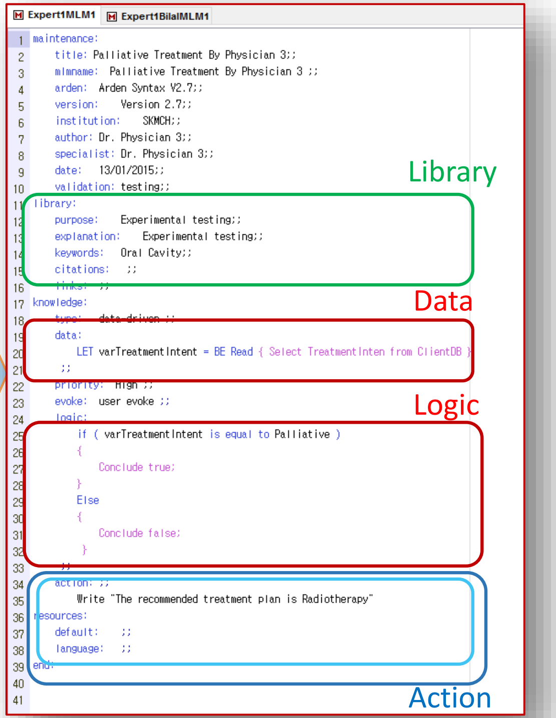
Sol 1-A

PICO Compliant Question Construction Example (8/8)

CDSS Recommendation

Surgery is recommended Treatment

Medical Logic Module



PICO Compliant Questions

P I C O
X

Treatment Intent Radical AND Tumor Stage 1 AND Surgery AND Radiotherapy

Ubiquitous Computing Laboratory
Kyung Hee University, Korea

Why Solution 1-A is not sufficient?

PICO Compliant Questions P I C O

Treatment Intent Radical AND Tumor Stage 1 AND Surgery AND Radiotherapy

\+ Clinical Task (Therapy)

58

Search results
Items: 1 to 20 of 58
<< First < Prev Page 1 of 3 Next > Last >>

1. Radical cystectomy in the trea... invasive bladder cancer: long-term results in 1,054 patients.
Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, Skinner E, Bochner B, Thangathurai D, Mikhail M, Raghavan D, Skinner DG.
J Clin Oncol. 2001 Feb 1;19(3):666-75.
PMID: 11157016
Similar articles

2. Marriage and ethnicity predict treatment in localized prostate carcinoma.
Denberg TD, Beaty BL, Kim FJ, Steiner JF.
Cancer. 2005 May 1;103(9):1819-25.
PMID: 15795905 Free Article
Similar articles

3. Radical radiotherapy for stage I/II non-small cell lung cancer in patients not sufficiently fit for or declining surgery (medically inoperable).
Rowell NP, Williams CJ.
Cochrane Database Syst Rev. 2001;(1):CD002935. Review. Update in: Cochrane Database Syst Rev. 2001;(2):CD002935.
PMID: 11279780
Similar articles

4. Radical prostatectomy after radiation therapy for cancer of the prostate: feasibility and prognosis.
Rainwater LM, Zincke H.

23

Results: 5 of 23

Radical treatment of synchro...us oligometa...tatic non-small cell lung carcinoma (NSCLC): pa...ent out...omes and prognostic factors.
Griffioen GH, Toguri D, Dahele M, Warner A, de Haan PF, Rodrigues GB, Slotman BJ, Yaremko BP, Senan S, Palma DA.
Lung Cancer. 2013 Oct; 82(1):95-102. Epub 2013 Aug 6.

[Recent advances in the treatment of laryngeal and hypopharyngeal carcinoma].
Eckel HE.
HNO. 2012 Jan; 60(1):6-18.

Extra-pleural pneumonectomy versus no extra-pleural pneumonectomy for patients with malignant pleural mesothelioma: clinical outcomes of the Mesothelioma and Radical Surgery (MARS) randomised feasibility study.
Treasure T, Lang-Lazdunski L, Waller D, Bliss JM, Tan C, Entwisle J, Snee M, O'Brien M, Thomas G, Senan S, et al.
Lancet Oncol. 2011 Aug; 12(8):763-72. Epub 2011 Jun 30.

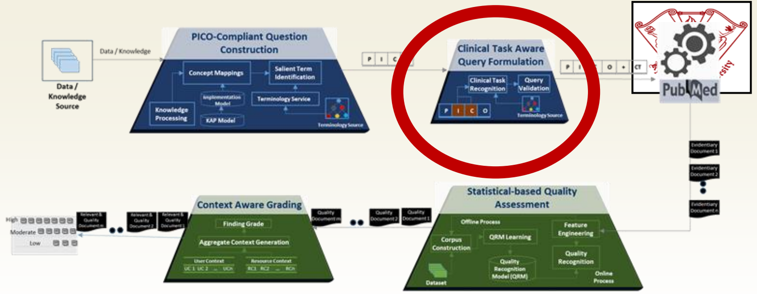
Effects of change in rectal cancer management on outcomes in British Columbia.
Phang PT, McGahan CE, McGregor G, MacFarlane JK, Brown CJ, Raval MJ, Cheifetz R, Hay JH.
Can J Surg. 2010 Aug; 53(4):225-31.

Sol 1-B

Clinical Task Aware Query Formulation

Sol 1-B

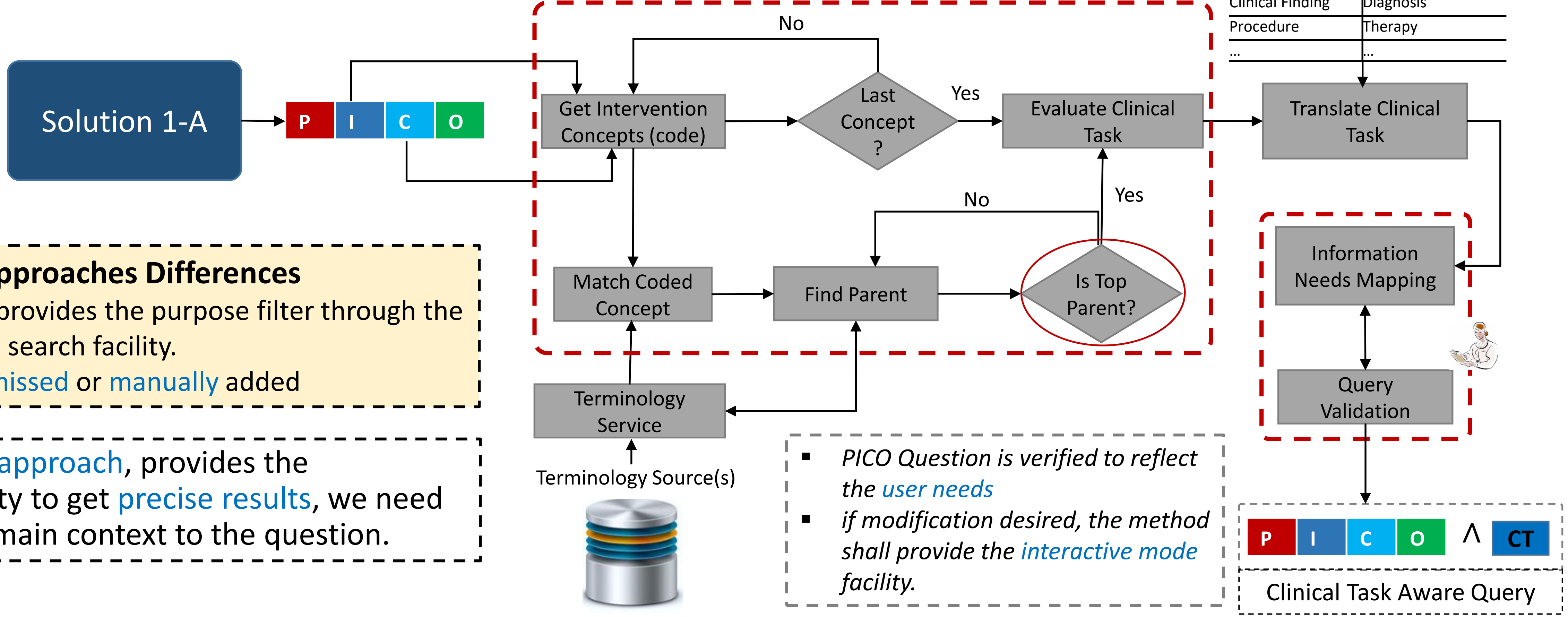
Clinical Task Aware Query Formulation (1/3)



- Clinical Task represents the user task.
- Technically, clinical task refers to the semantic category/group in vocabulary systems such as SNOMED CT, UMLS.

Translation Table

Semantic Category	Clinical Task
Clinical Finding	Diagnosis
Procedure	Therapy
…	…



Existing Approaches Differences

- PubMed provides the purpose filter through the advanced search facility.
- Majorly missed or manually added

Proposed approach, provides the opportunity to get precise results, we need to add domain context to the question.

- *PICO Question is verified to reflect the user needs*
- *if modification desired, the method shall provide the interactive mode facility.*

[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

Sol 1-B

Clinical Task Aware Query Formulation (2/3)

Problem | Procedure

Clinical Task Recognition (CTR) and Translation Algorithm

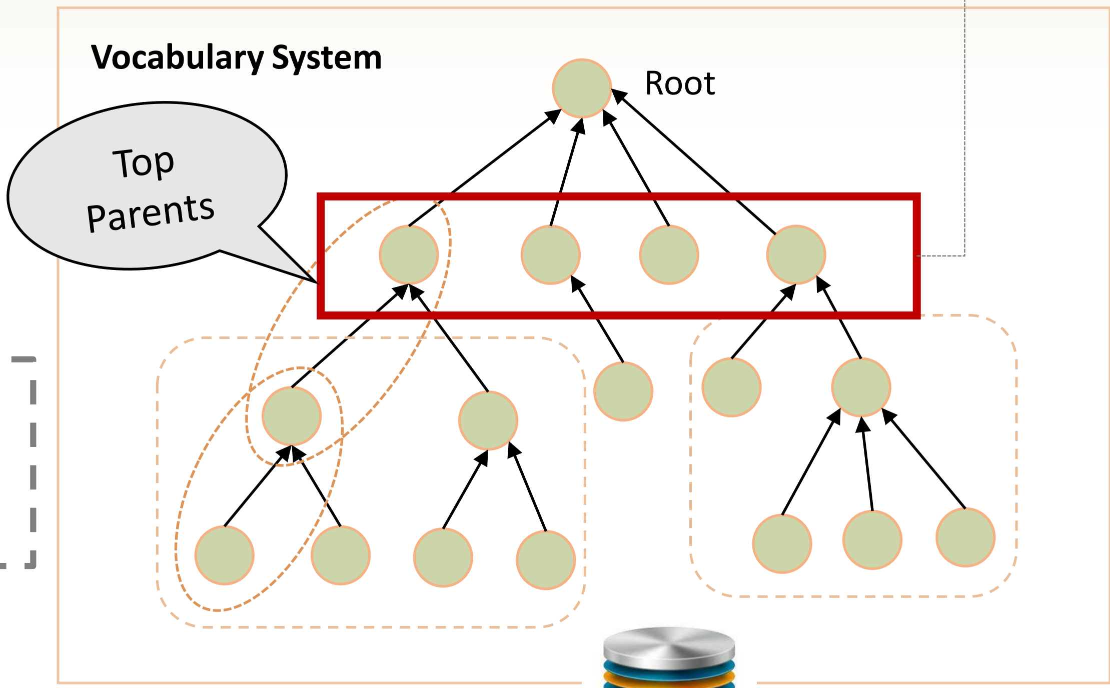
Ultimate Parent finding

```
Begin
inputs:  I, C – {t1, t2, …, tk }; // the list of intervention terms
output:  TP - {p1, p2, …, pk}; // the list of top parent concepts

    for each c in I
        parent = STS.findParent(c)
        if (parent == top parent)
            AddParent(PC, parent)
        else
            call STS.findParent(parent)
        end if
    end for
    return PC;
```

STS is a Standard Terminology Service

- STS is available for UMLS and SNOMED CT
- **UMLS** → MetaMap and SemMap services
- **SNOMED CT** → IHTSDO service



Clinical Task finding

```
for each c in PC
Begin
    cDist[] = countDisctinct(c)
End
QT = cDist with highest value
return QT
End
```

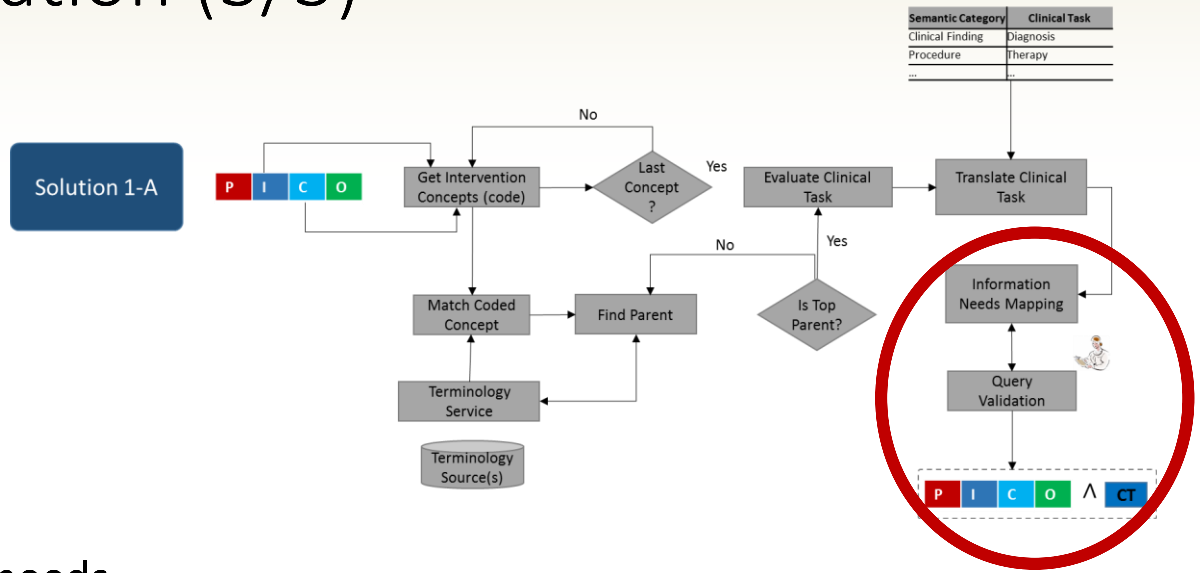
Clinical Task Translation (CTT)

```
Begin
inputs:  β; //where  β represents the query type
output:  β' // where  β' is translated term of parent concept  β
    if (  β = "clinical fin
        β'←"Diagonis";
    elseif ( β = "procedure")
        β'←"Therapy";
    endif
        return   β';
End
```

Translation is required by the target search engine to recognize the correct document type.

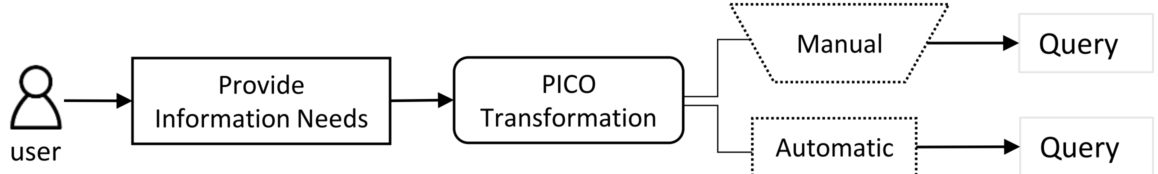
Sol 1-B

Clinical Task Aware Query Formulation (3/3)



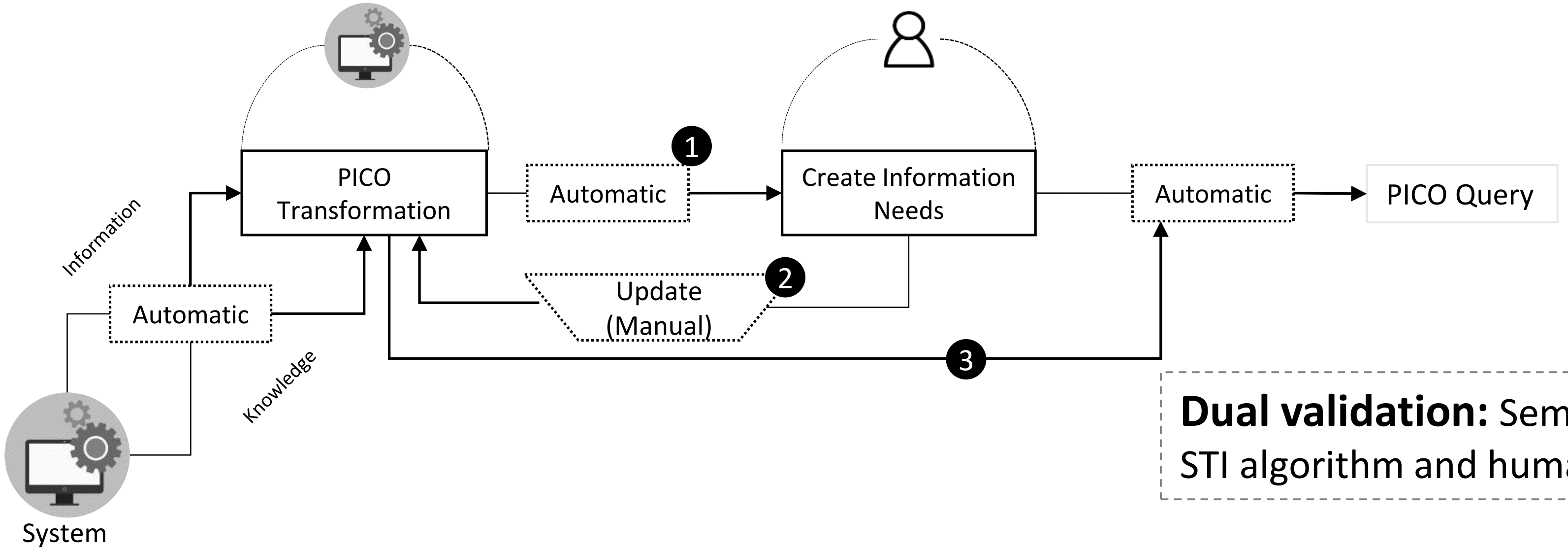
Existing Approach

Approach A: From information needs to PICO



Proposed Approach

Approach B: From information/knowledge to PICO to information needs



Information need represents the intended question user is interested (usually in natural language format).

Query is the search engine acceptable representation.

Dual validation: Semantic group based validation using extended STI algorithm and human verification of information needs.

[Afzal2015] Afzal, Muhammad, et al. "Knowledge-Based Query Construction Using the CDSS Knowledge Base for Efficient Evidence Retrieval." Sensors 15.9 (2015): 21294-21314.

Sol 1

Experiments Results

Document Retrieval Accuracy

- **Evaluation Criteria** [Wilczynski2005]
  - **P10, MP, TDDR, MRR**

  P10: Precision at 10 retrieved documents

$$P10 = \frac{a}{a+b}$$

$\boldsymbol{a}$ = true positives, articles found by the search term meet the criteria

$\boldsymbol{b}$ = false positives, articles found by the search term do not meet the criteria

$$P10 = \frac{(Precision \times 10)}{10}$$ …Scaled P10 when no. of docs < 10

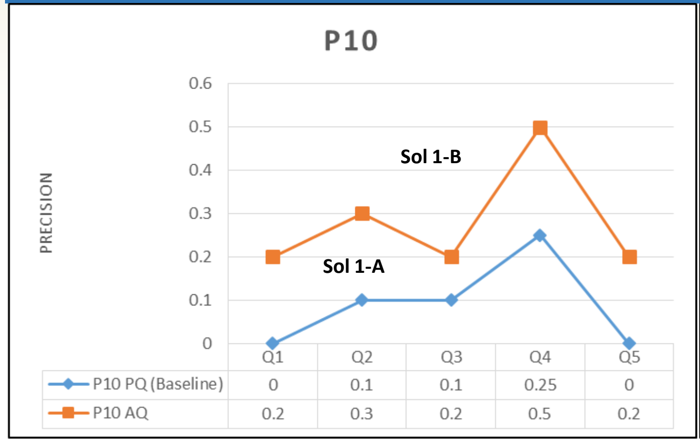
MP: Mean Precision for all queries

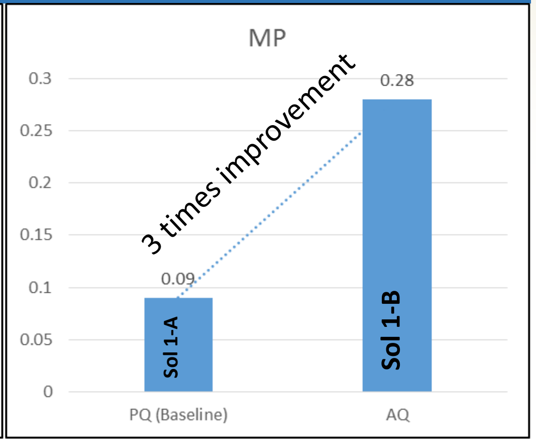
TDDR: Total Document Reciprocal Rank

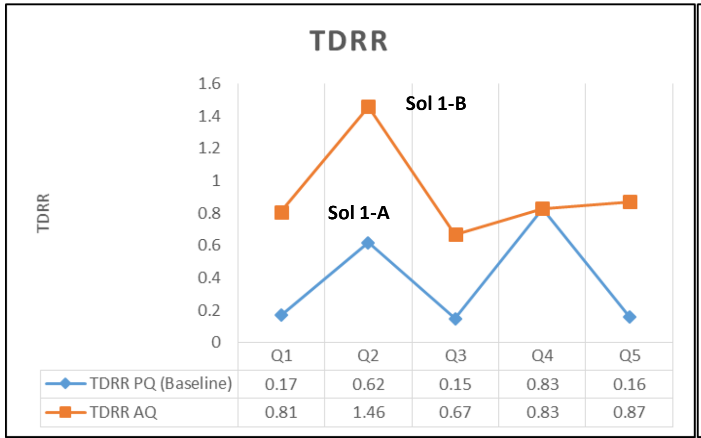
MRR: Mean Reciprocal Rank for all queries

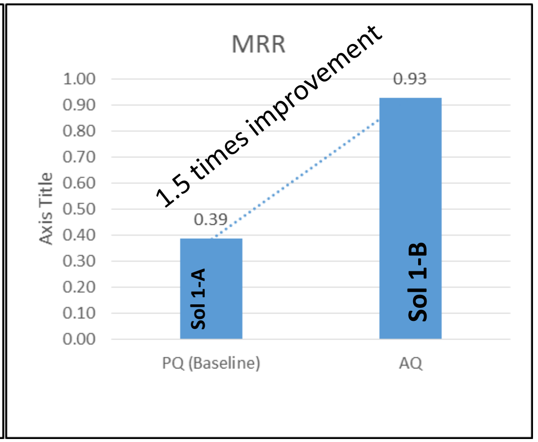
- **Experimental Setup**
  - PubMed search engine
  - Medline database is used for searching
- **Dataset**
  - **7 MLMs** from public domain [Maq2015]
  - **3 MLMs** are additionally created by domain experts
  - **15 queries** derived from selected MLMs

Results (Comparison between Sol-A and Sol-B)









Evaluations:

- P10 for AQ **(Sol 1-B)** performance is found better than **PQ (Sol 1-A)** for all the queries.
- MP for AQ showed 3 times improved performance to PQ.
- PQ performed poorly in all cases for TDRR except the fourth query
- AQ showed around 1.5 times improved performance than PQ in MRR.

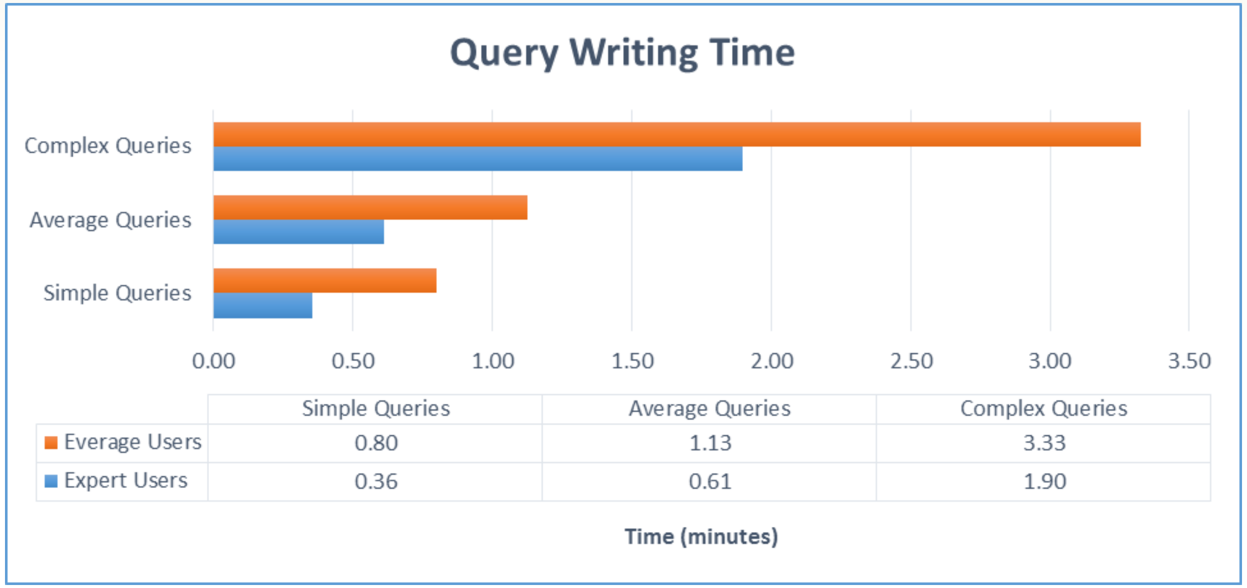
Sol 1

Experiments Results

Document Retrieval Accuracy

- **Evaluation Criteria**
  - **Query writing time (minutes)**
- **Experiment environment**
  - Three type of queries
    - simple (consisting of <3 terms)
    - average (consisting of between 4 and 8 terms)
    - complex (consisting of >8 terms)
  - Experiment is performed by writing the auto constructed queries manually in PubMed browser.
  - Two type of users: average and expert.

Results



	Simple Queries	Average Queries	Complex Queries
Everage Users	0.80	1.13	3.33
Expert Users	0.36	0.61	1.90

Evaluations:

- Overall, the automated query construction process saved on the average about 0.90 minutes for all quarries.
- For the expert users, it saved 1.75 minutes on the average for all queries.

Solution 1 Summary

A PICO Compliant Question Construction

B Clinical Task Aware Query Formulation

Contributions

- Mapping Model (KAP) Construction
- Salient Term Identification
- Clinical Task Recognition
- PICO Query Validation

Need for Solution 2

Query: "oral cavity cancer"

Document 1

Format: Abstract ▾ Send to ▾

Laryngoscope. 2015 Aug;125(8):1869-73. doi: 10.1002/lary.25328. Epub 2015 Jun 9.

Complications and mortality following surgery for oral cavity cancer: analysis of 408 cases.

Schwam ZG[1], Sosa JA[2,3,4], Roman S[2], Judson BL[1].

⊕ Author information

Abstract

OBJECTIVES: To analyze the postoperative complications and mortality for oral cavity cancers, their time course, and to identify modifiable risk factors associated with their occurrence.

STUDY DESIGN: Retrospective cohort study.

METHODS: Patients undergoing surgery for oral cavity cancer were identified in the American College of Surgeons National Surgical Quality Improvement Program Participant Use Data File (2005-2010). Overall and disease-specific complication and mortality data were analyzed using chi-square and multivariate regression analysis.

RESULTS: There were 408 cases identified. The overall 30-day complication and mortality rates were 20.3% and 1.0%, respectively. The most common adverse events were reoperation (9.6%), infectious (6.6%), and respiratory (5.1%) complications. Twenty patients (4.9%) experienced postdischarge complications. Fifty-two percent of postdischarge wound dehiscences and 67% of postdischarge surgical-site infections occurred by postdischarge day 7, and 91% of all postdischarge complications occurred by postdischarge day 14. Smoking was independently associated with respiratory (odds ratio [OR] 3.59, P = .008) and surgical site complications (OR 5.13, P = .004). Neck dissection was independently associated with respiratory (OR 6.17, P = .001), surgical site (OR 6.30, P = .003), and infectious (OR 3.83, P = .003) complications.

CONCLUSION: Current smokers and those undergoing neck dissection are at high risk of postoperative complications after oral cavity cancer surgery. Less than 5% of patients experienced postdischarge complications, nearly all of which occurred by postdischarge day 14. Most early postdischarge complications occurred at the surgical site. In order to mitigate postdischarge complications and their sequelae, early clinical follow-up should be sought for high-risk patients.

LEVEL OF EVIDENCE: 4.

© 2015 The American Laryngological, Rhinological and Otological Society, Inc.

KEYWORDS: Head and neck cancer; NSQIP; oral cavity cancer

PMID: 26063059 DOI: 10.1002/lary.25328

[PubMed - indexed for MEDLINE]

Document 2

BMC Cancer. 2015 Oct 31;15:827. doi: 10.1186/s12885-015-1841-5.

Population attributable risks of oral cavity cancer to behavioral and medical risk factors in France: results of a large population-based case-control study, the ICARE study.

Radoï L[1,2], Menvielle G[3,4], Cyr D[5,6], Lapôtre-Ledoux B[7], Stücker I[8,9], Luce D[10,11]; ICARE Study Group.

⊕ Collaborators (27)

⊕ Author information

Abstract

BACKGROUND: Population attributable risks (PARs) are useful tool to estimate the burden of risk factors in cancer incidence. Few studies estimated the PARs of oral cavity cancer to tobacco smoking alone, alcohol drinking alone and their joint consumption but none performed analysis stratified by subsite, gender or age. Among the suspected risk factors of oral cavity cancer, only PAR to a family history of head and neck cancer was reported in two studies. The purpose of this study was to estimate in France the PARs of oral cavity cancer to several recognized and suspected risk factors, overall and by subsite, gender and age.

METHODS: We analysed data from 6[illegible]9 oral cavity cancer cases and 3481 controls included in a population-based case-control study, the ICARE study. Unconditional logistic regression models were used to estimate odds ratios (ORs), PARs and 95% confidence intervals (95% CI).

RESULTS: The PARs were 0.3% (95% CI -3.9%; +3.9%) for alcohol alone, 12.7% (6.9%-18.0%) for tobacco alone and 69.9% (64.4%-74.7%) for their joint consumption. PAR to combined alcohol and tobacco consumption was 74% (66.5%-79.9%) in men and 45.4% (32.7%-55.6%) in women. Among suspected risk factors, body mass index 2 years before the interview <25 kg.m(-2), never tea drinking and family history of head and neck cancer explained 35.3% (25.7%-43.6%), 30.3% (14.4%-43.3%) and 5.8% (0.6%-10.8%) of cancer burden, respectively. About 93% (88.3%-95.6%) of oral cavity cancers were explained by all risk factors, 94.3% (88.4%-97.2%) in men and only 74.1% (47.0%-87.3%) in women.

CONCLUSION: Our study emphasizes the role of combined tobacco and alcohol consumption in the oral cavity cancer burden in France and gives an indication of the proportion of cases attributable to other risk factors. Most of oral cavity cancers are attributable to concurrent smoking and drinking and would be potentially preventable through smoking or drinking cessation. If the majority of cases are explained by recognized or suspected risk factors in men, a substantial number of cancers in women are probably due to still unexplored factors that remain to be clarified by future studies.

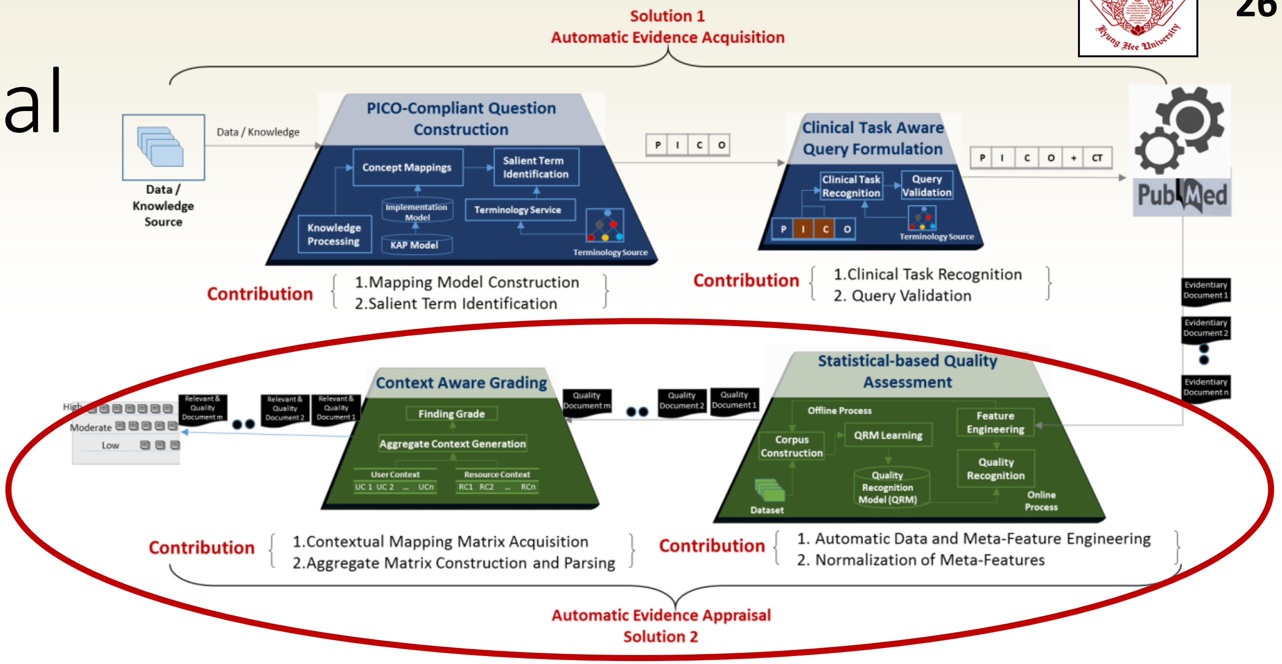
PMID: 26520570 PMCID: PMC4628276 DOI: 10.1186/s12885-015-1841-5

[PubMed - indexed for MEDLINE] Free PMC Article

- Based on keyword matching technique, both documents will be retrieved and both will have equal importance because they there exist one match in each with the query.
- A well-built query can only provides relevance. It cannot guarantees the quality of the contents.

Sol 2

Automatic Evidence Appraisal



Existing Approaches Differences

- Insufficient and unreliable datasets
- Manual engineering of meta-features
- Non-normalized meta-features
- Based only on resource context

Disadvantages

- Time consuming for experienced physicians
- Hard for inexperienced users (nurse, patients)
- Comparatively less accurate

Solution 2 provides methods to identify quality evidences on the basis of a statistical model that uses.

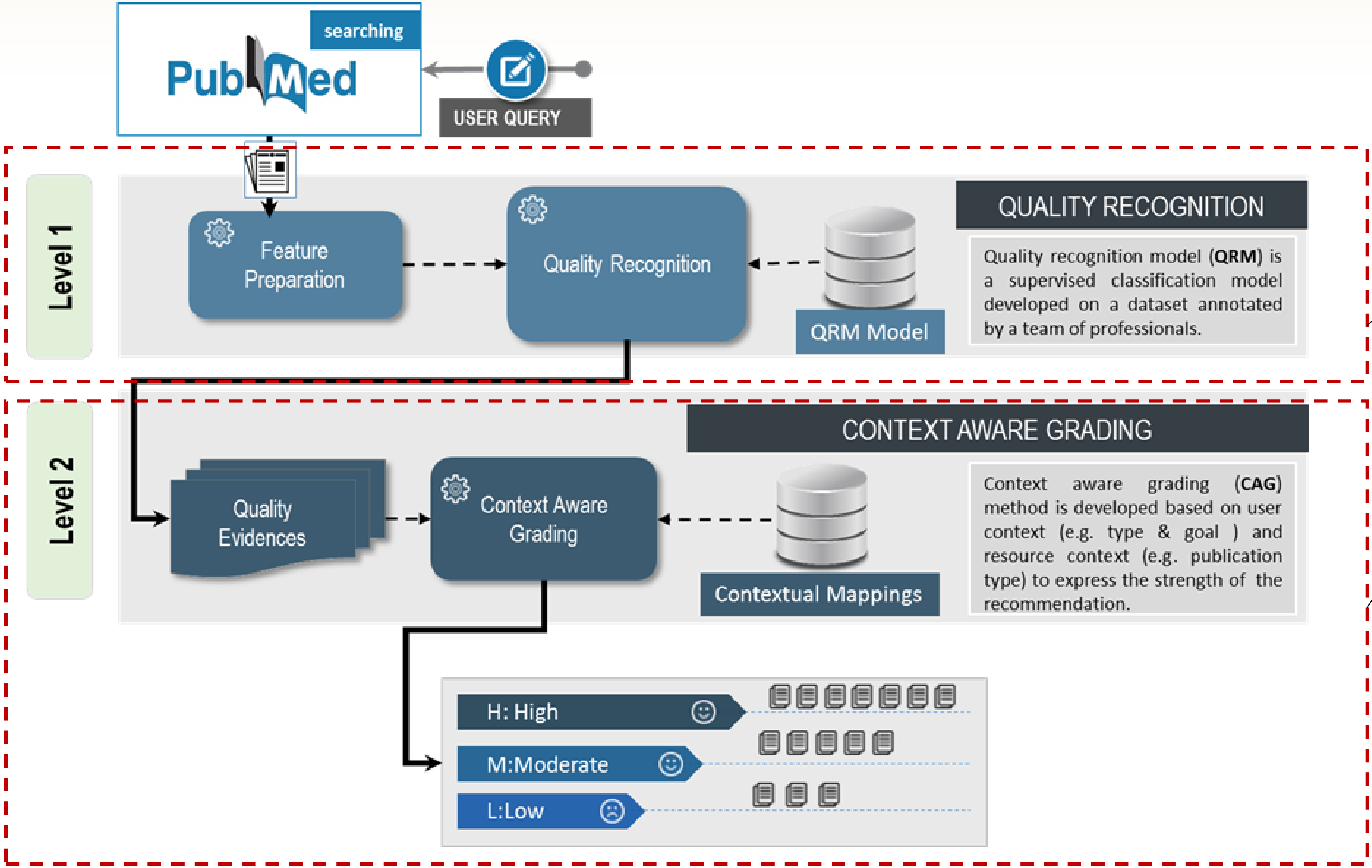
- a dataset annotated by a team of expert
- An automatic method for meta-feature engineering
- User and resource aggregate contextual grading

Quality Evidence Definition:

- An evidence is considered as scientifically rigorous if its analysis is consistent with the study design [21]

Sol 2

Abstract Methodology



Sol 2-A

Statistical-based Quality Assessment

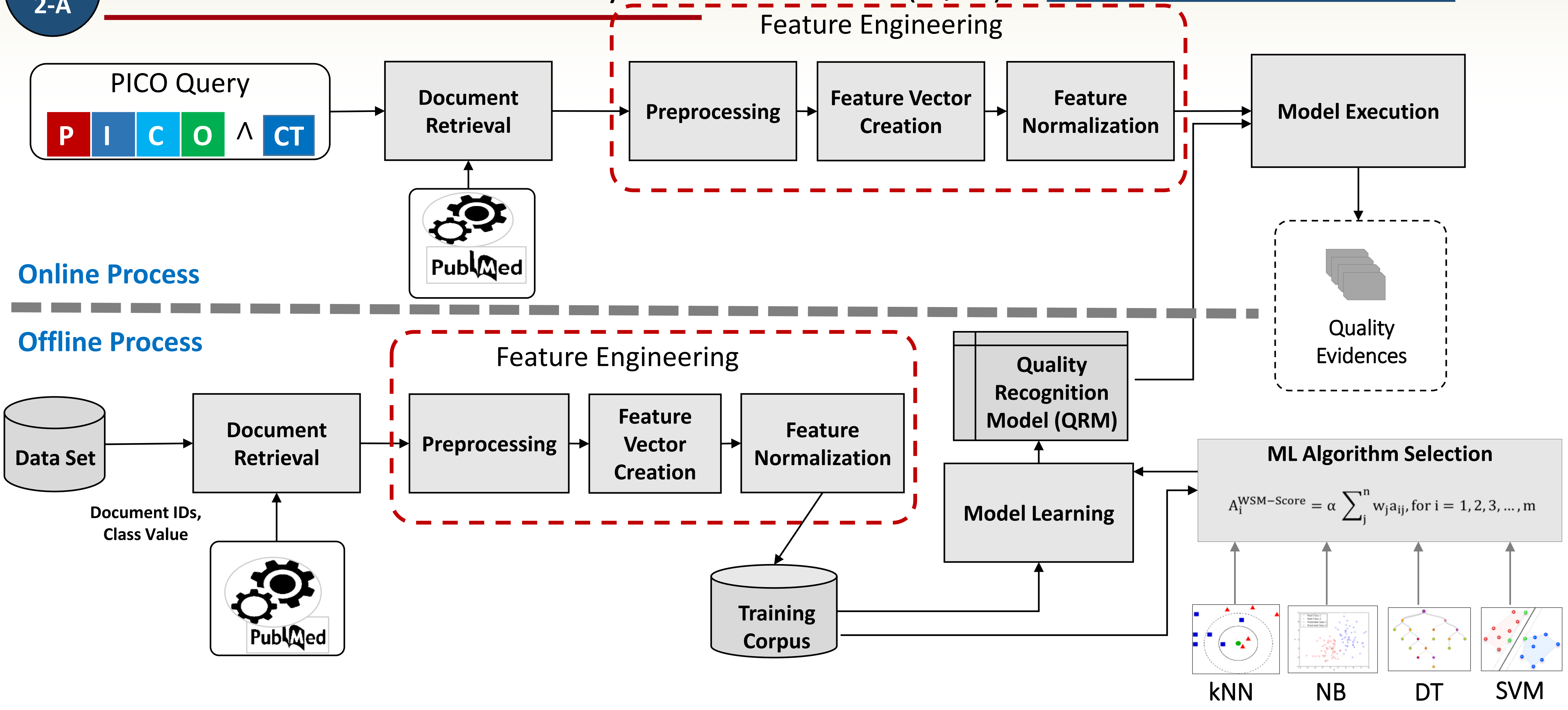
Sol 2-B

Context Aware Grading

Sol 2-A

Statistical-based Quality Assessment (1/5)

Quality Recognition Model



Sol 2-A

Statistical-based Quality Assessment (2/5)

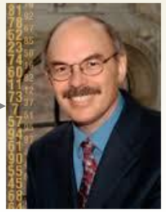
Dataset Selection

Dataset Issues

- For statistical approaches, the most crucial part is the selection of dataset.
- Small dataset are not trustable.
- Annotation from the domain experts with acceptable mutual agreement.
- Availability and reliability

Dataset in Proposed Method

- A dataset that was manually created by a team of experts.
- An agreement (authorship inclusion) signed with R.B. Haynes.



R.B Haynes
Chief, Health Information Reserach Unit, McMaster University
Editor, ACP (American College of Physician) Journal Club

Characteristics of dataset

Sno.	PubMedId	Format	HHC	Purpose	Rigor
1	10601047	O	TRUE	P	FALSE
2	10601048	O	TRUE	P	FALSE
3	10601049	O	TRUE	SE	FALSE

50593	10601388	GM	FALSE		FALSE
50594	10601389	GM	FALSE		FALSE

Format
- **O**: Original study
- **R**: Review
- **GM**: General and miscellaneous articles
- **CR**: Case report

HCC (Of interest to the health care of humans)
- True
- False

Purpose
- **Tr**: Treatment
- **D**: Diagnosis
- **P**: Prognosis
- **E**: Etiology
- ●●

Rigor (Methodological Rigorousness)
- True
- False

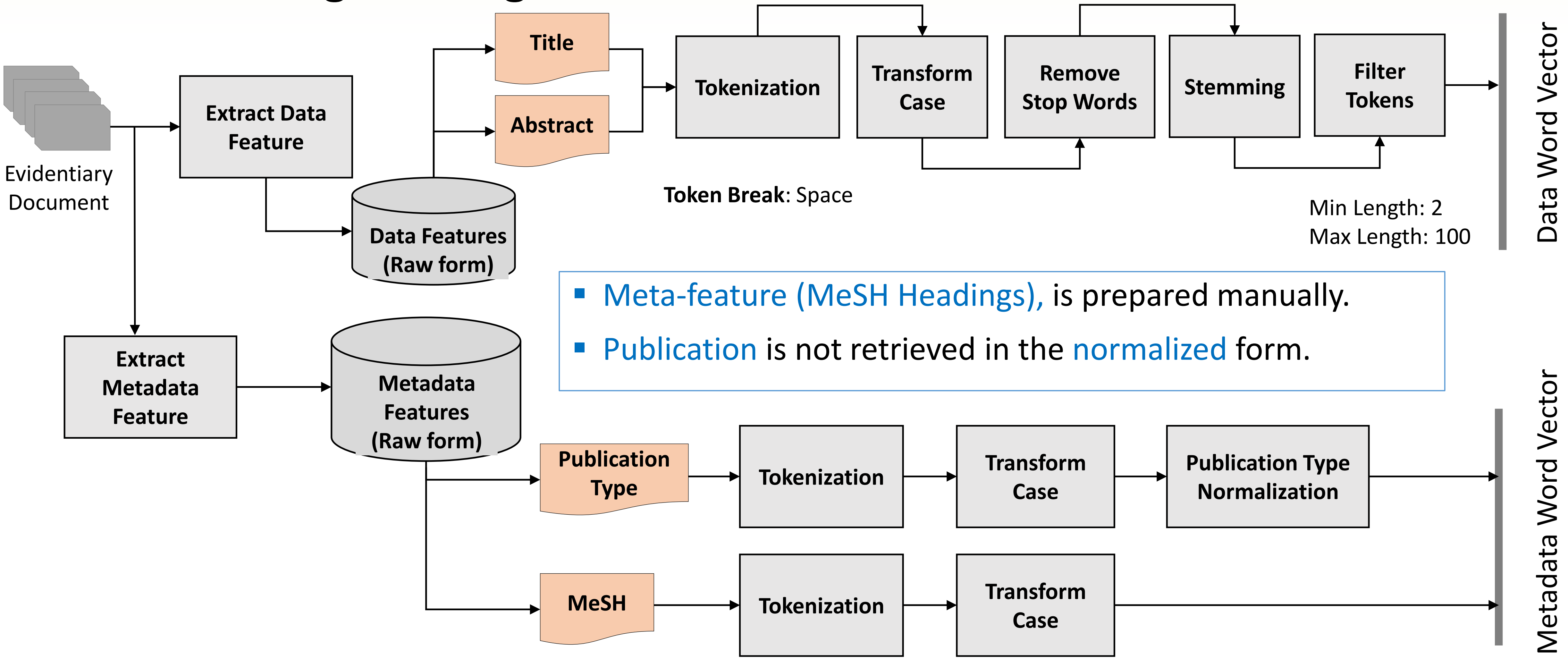
PubMedId shows the unique identifier of ACP Journal Club article

[AfzalCAG2016] Afzal, Muhammad, et al. "Context Aware Grading of Quality Evidences for Evidence-based Decision Making"" Health Informatics Journal (SAGE) (Minor Revision)

Sol 2-A

Statistical-based Quality Assessment (3/5)

Feature Engineering



- Meta-feature (MeSH Headings), is prepared manually.
- Publication is not retrieved in the normalized form.

Sol 2-A

Statistical-based Quality Assessment (4/5)

Feature Engineering

Feature Extraction Algorithm

```
Begin

    inputs: PMIDs – {id1, id2, ..., idn}; //list of PubMed ids of training dataset

    output: F – {f1, f2, f3, f4}  /* where f1 = title, f2 = abstract, f3 =
                                     MeSH, and f4 = publication type */

    Let ePostResultRef is the reference to the database of uploaded IDs
    ePostResultRef ← ePost(PMIDs); //upload the PMIDs list to PubMed database
    eFetchResult ← eFetch(ePostResultRef); //download the documents

    for i = 0 to eFetchResult.count – 1
        f1 ← i.title;
        F.add(f1);
        f2 ← i.abstractText;
        F.add(f2);
        f3 ← "";
        for j = 0 to i.MeSHHeading.count – 1
            f3 ← f3 + i.MeSHHeading;
        endfor
        F.add(f3);
        f4 ← "";
        for m = 0 to i.publicationtype.count – 1
            f4 ← f4 + "," + i.publicationtype;
        endfor
        F.add(f4);
    endfor

    Return F;

End
```

Publication Type Standardization Algorithm

```
Begin

    inputs: A – {a1, a2, ..., an}; //the list of articles
    output: A' – {a1, a2, ..., an}; // the list of articles with standardized publication type

    Let;
        pt represents publication type;
        rank represents the rank of pt;
        tempRank = 0; // holds the previous rank temporarily for comparison
        spt represents the standardized publication type;
    for each a in A
        do
            pt ← a.getPublicationType();
            rank ← getRank(pt, R); //where R is the rank table for publication types.
                if (rank > tempRank)
                tempRank ← rank;
                spt ← pt;
            endif
        while (a.getPublicationType exists)
         a.PublicationType ← spt;
        A'.add(a);
    endfor
    return A';
End
```

Clinical Task	Resource Type
Diagnosis	Prospective, blind comparison to a gold standard or cross-sectional
Therapy	randomized controlled trial > cohort study
Prognosis	cohort study > case control > case series
Harm/Etiology	cohort > case control > case series

Publication Type	Rank
Meta-analysis of RCTs	1
Systematic Review of RCTs	2
Randomized Controlled Trial (RCT)	3
Meta-analysis of CTs	4
Systematic Review of CTs	5
...	...

[AfzalCAG2016] Afzal, Muhammad, et al. "Context Aware Grading of Quality Evidences for Evidence-based Decision Making" Health Informatics Journal (SAGE) (Minor Revision)

Sol 2-A

Statistical-based Quality Assessment (5/5)

Machine Learning Method Selection

- A set of methods have been tried.
- DT, SVM, NB, and kNN ranked on the top

	Training			Testing				
Algorithm/Criteria	F-Measure	Accuracy	AUC	F-Measure	Accuracy	AUC	Sum Score	Scaled Ranking
SVM	0.849	0.771	0.807	0.870	0.785	0.735	4.818	0.80
DT	0.914	0.883	0.969	0.289	0.316	0.762	4.134	0.69
NB	0.835	0.764	0.752	0.721	0.602	0.548	4.223	0.70
kNN	0.812	0.707	0.782	0.847	0.752	0.777	4.678	0.78

$$A_i^{WSM-Score} = \alpha \sum_{j=1}^{m} w_j\, a_{ij}, \; for\ i = 1, 2, 3, \ldots, m$$

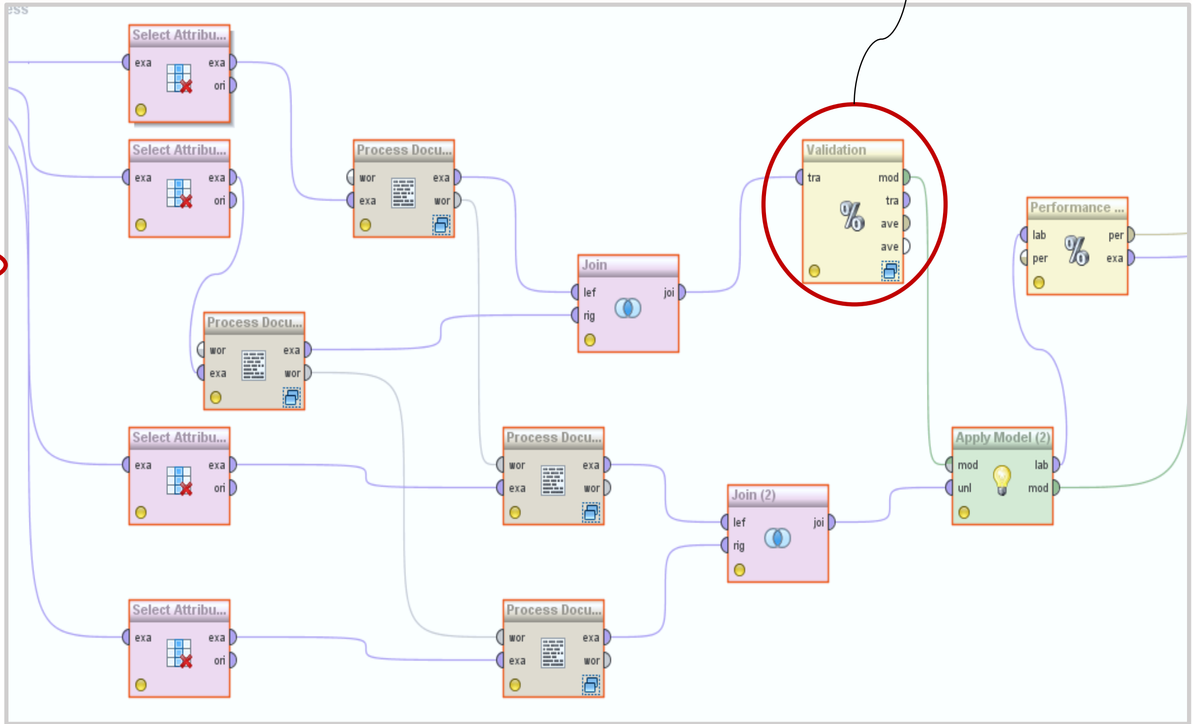
SVM Parameter Setting

kernel cache	200
C	0.0
convergence epsilon	0.001
max iterations	100000
☑ scale	
L pos	1.0
L neg	1.0
epsilon	0.0

- Complex cost parameter C values less than 0.0 showed similar results to C = 0.0.
- Similarly, values greater than 0.1 produces almost similar results to C = 0.1.
- The kernel cache value is set to 200 and maximum iterations is set to 100000.
- Finally we were left with C = 0.0 and C = 0.1 to choose from however, C = 0.0 for our experiment produced better results as compared to C = 0.1.

Quality Recognition Model

10-fold cross validation



QRM is a binary classification model used to predict methodologically rigorous articles (high quality evidences).

Sol 2-B

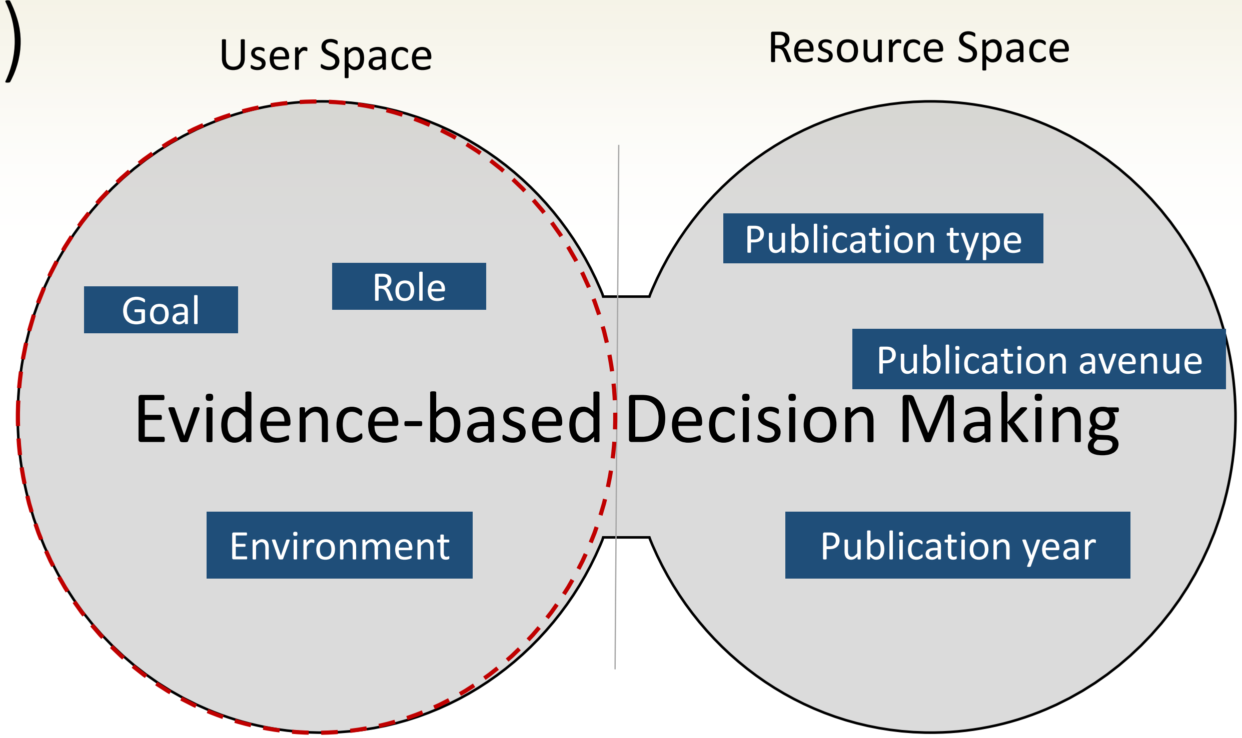
Context Aware Grading (1/3)

Existing approaches

- Rely mostly on resource context to grade evidences [Sarker2015].
  - Publication type
  - Publication avenue
  - Publication

Issue:

- Missing to reflect the stakeholder (user) aspects
  - Role, Goal, Environment



Proposed Approach

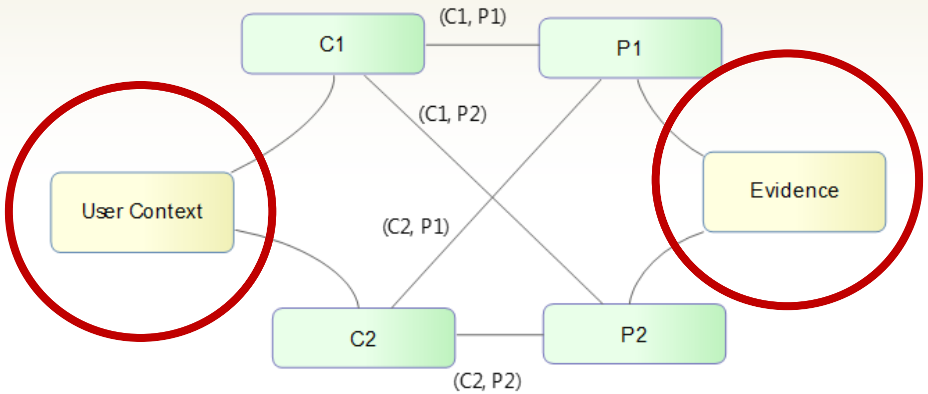
- Add user context with resource context (SORT).
- Based on PARIHS Framework [PARiHS2004] and Verbert Context Framework [Verbert2012]

- SORT (Strength of Recommendation Taxonomy)
  - is a well-recognized grading system in EBM community [Ebell2004].

Sol 2-B

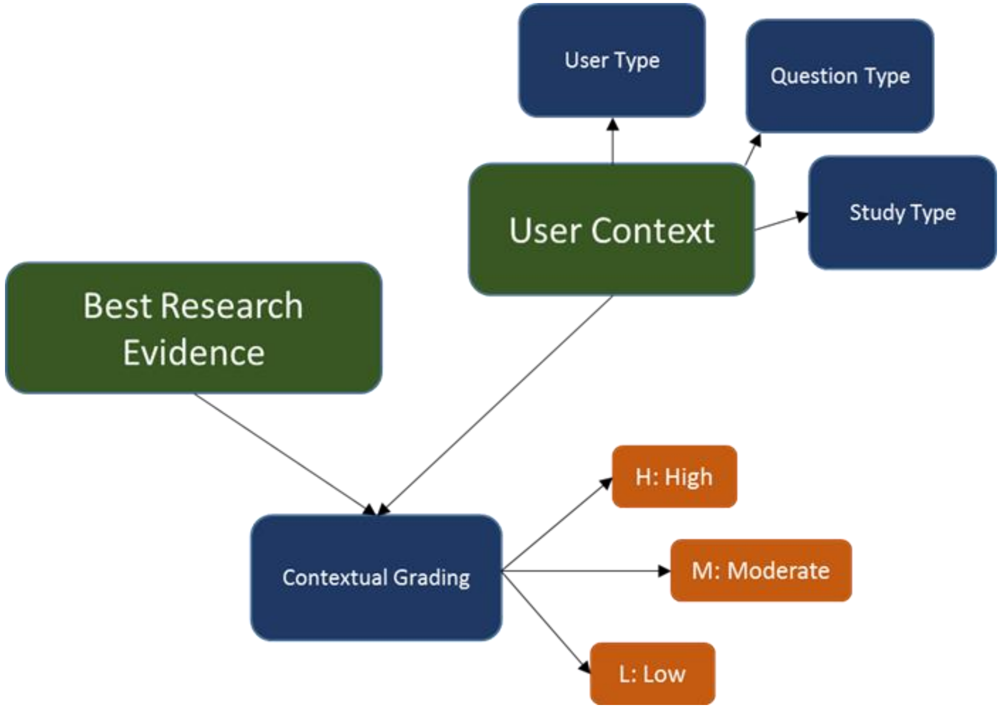
Context Aware Grading (2/3)

User Aware SORT-Based Evidence Grading



Contextual mapping tables

- Literature-based analysis
  - [Ebell2004], [Wilczynski2005], [WG2004].
- Expert-based
  - Questionnaire filled from the domain experts



Grade value population for an evidence with respect to contexts

Context\Evidence	P1	P2	…	Pn
C1	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)
C2	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)
…	…	…	…	…
Cn	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)

Context Aware SORT Grading

Sol 2-B

Context Aware Grading (3/3)

Contextual Evidence Grading Algorithm

```
Begin
input:  E<e> //the list of rigor evidences
output:  G<e,g> // where g represents the grades h, m, l, u
Let;
        C <c> //current context
        P <p> //properties of E
        G <g> // grade values
        for each e in E
                for each p in P
                    for each c in C
                        grade ← computeGrade(p,c)
                        G.add(grade)
                    endfor
                endfor
                finalGrade ← getHighestGrade(G)
                GE ← addGrade(e,finalGrade);
        endfor
        return GE;
End
```

User context is captured from the source system

User Context	Resource Context
Diagnosis	prospective, blind comparison to a gold standard or cross-sectional
Therapy	randomized controlled trial > cohort study
Prognosis	cohort study > case control > case series
Harm/Etiology	cohort > case control > case series

Resource context is created from the PubMed retrieved documents

Context\Evidence	P1	P2	…	Pn	
C1	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)	
C2	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)	
…	…	…	…	…	
Cn	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)	
Aggregate Contextual Grade Values	(H or M or L or U)	(H or M or L or U)	…	(H or M or L or U)	(H or M or L or U)

Aggregate Contextual Grade Vector

Final Grade Value

$$\mathit{Final\ Grade}\ \text{Value} = \text{Max}\left(\begin{cases} \text{HCount} = \sum_{i=1}^{n} H_i \\ \text{MCount} = \sum_{i=1}^{n} M_i \\ \text{LCount} = \sum_{i=1}^{n} L_i \\ \text{UCount} = \sum_{i=1}^{n} U_i \end{cases}\right)$$

[AfzalCAG2016] Afzal, Muhammad, et al. "Context Aware Grading of Quality Evidences for Evidence-based Decision Making" Health Informatics Journal (SAGE) (Minor Revision)

Sol 2

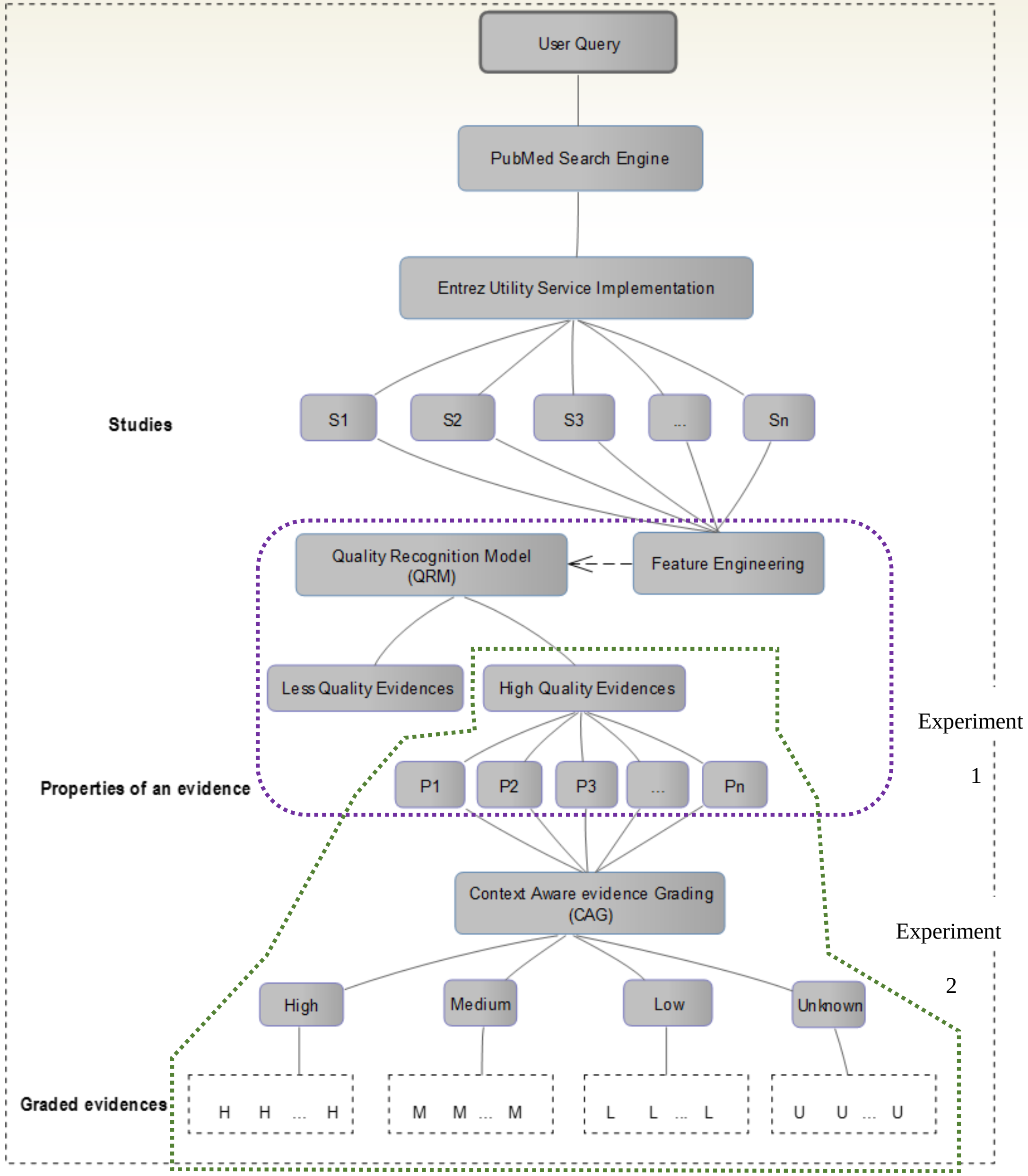
Experiment Setup

- **Experiments**
  - **Experiment 1:** QRM Performance
  - **Experiment 2:** CAG Performance

- **Experimental Setup**
  - RapidMiner Studio Basic 6.5.002
  - Windows 10, RAM 4GB
  - Search Engine: PubMed

- **Evaluations**
  - Statistical Evaluation (Recall, Precision, F-Measure, and Accuracy)
  - Human Evaluation (two oncologists as domain experts)

- **Dataset**
  - Training Dataset: 5682 Therapy related Medline articles
  - Development Test Dataset: 1300 articles



Sol 2

Experimental Results

Experiment 1: QRM Performance (SVM-Based Model)

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

Where;

$TP = True\ Positive,$

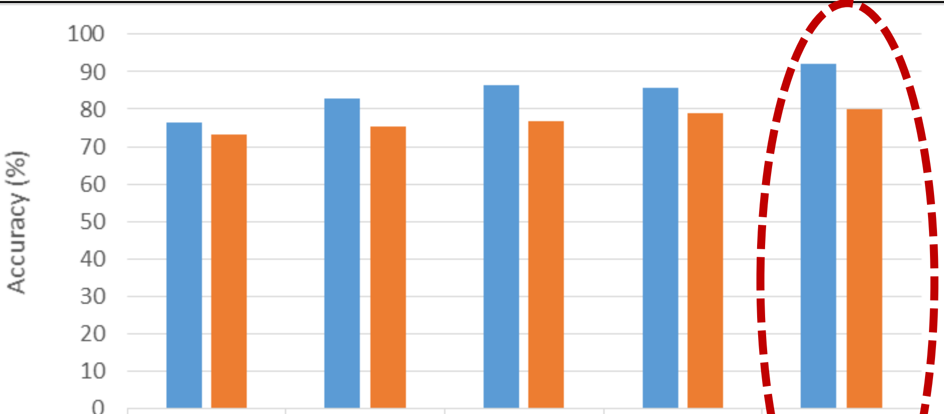
$FP = False\ Positive,$

$FN = False\ Negative, and$

$TN = True\ Negative$

SPT: Standard Publication Type
MeSH: Medical Subject Headings

Quality Recognition Model (QRM) Performance



	Title	Abstract	MeSH	SPT	All
Training	76.28	82.81	86.4	85.71	92.14
Testing	73.31	75.46	76.9	79	80.15

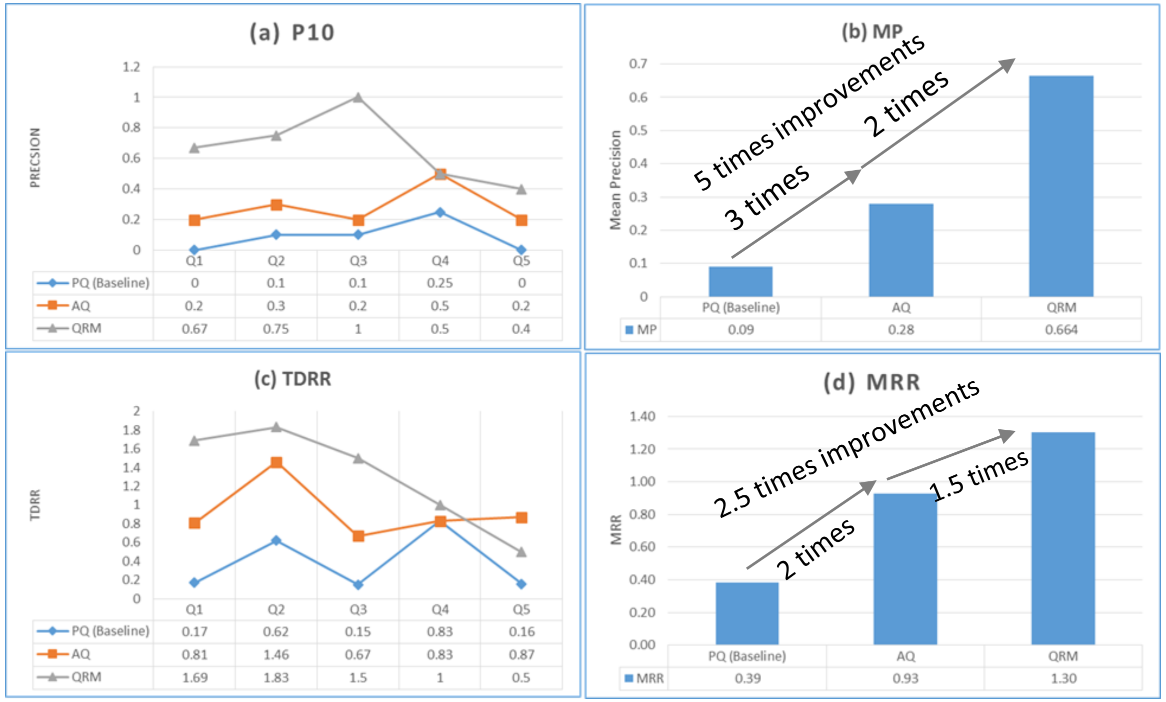
- Title feature remains the lowest in both training and testing cases and abstract feature remains second lowest.

- QRM performed exceptionally well on the combination of all features with 92.14% accuracy on training and 80.15% on testing dataset.

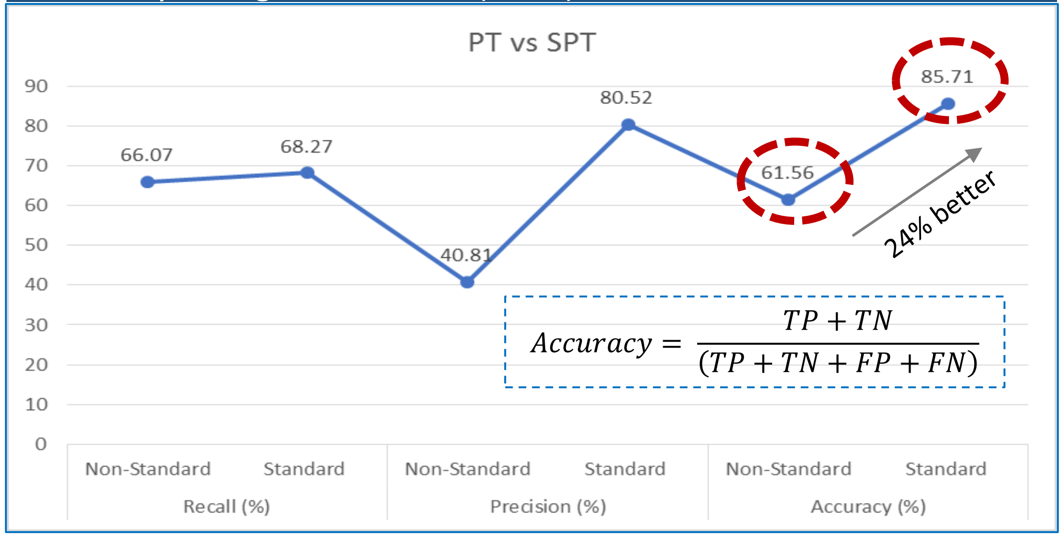
Sol 2

Experimental Results

- **Evaluation Criteria**
  - **P10, MP, TDDR, MRR**
    - ✓ **P10: Precision at 10 retrieved documents**
    - ✓ **MP: Mean Precision for all queries**
    - ✓ **TDDR: Total Document Reciprocal Rank**
    - ✓ **MRR: Mean Reciprocal Rank for all queries**
- **Experimental Setup**
  - **PubMed search engine**



Quality Recognition Model (QRM) Performance on PT and SPT



$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

Comparison with existing approaches in quality recognition

System	Accuracy
[Sarker2015]	76.38 %
Proposed System	80.85 %

About 4% Better

System	F-Measure
[Kilicoglu2009]	65.90 %
Proposed System	71.60 %

About 6% Better

[AfzalCAG2016] Afzal, Muhammad, et al. "Context Aware Grading of Quality Evidences for Evidence-based Decision Making" Health Informatics Journal (SAGE) (Minor Revision)

Sol 2

Experimental Results

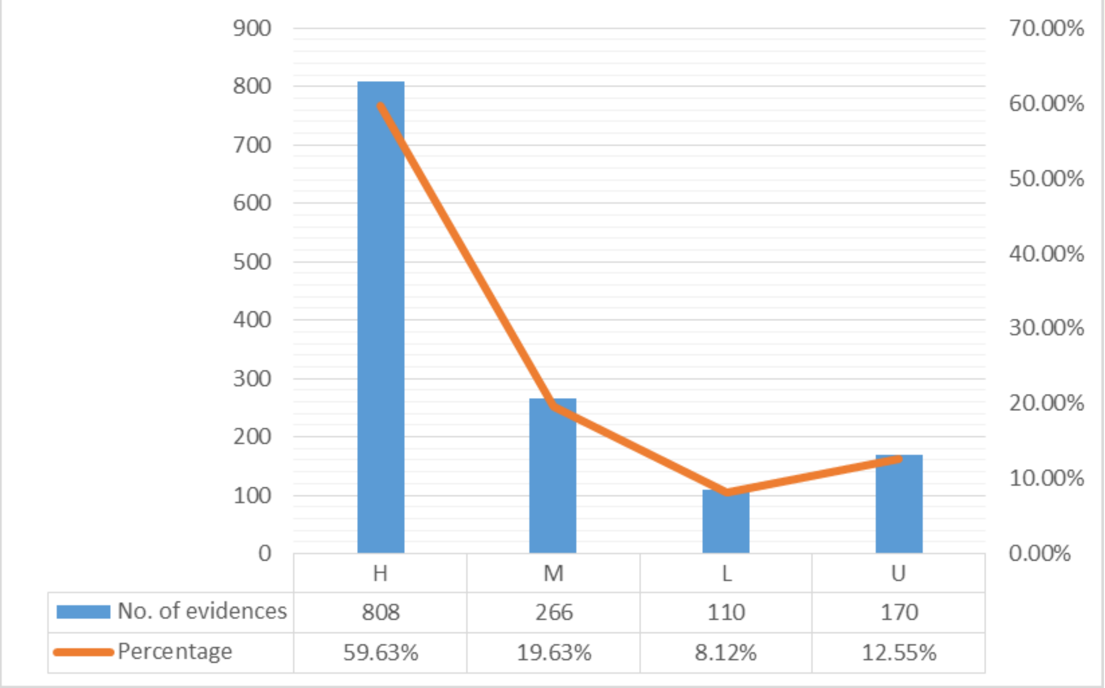
Experiment 2: Context Aware Grading (CAG) Performance

Out of 1355 documents, about 60% evidences are graded as H which means highly beneficial for the physician.

This contextual grading helps to re-rank the documents by bringing H evidences on the top followed by M.

For the given study, user context was Treatment as a user task and resource context was Publication Type.

CAG Performance to grade evidences on the basis of context

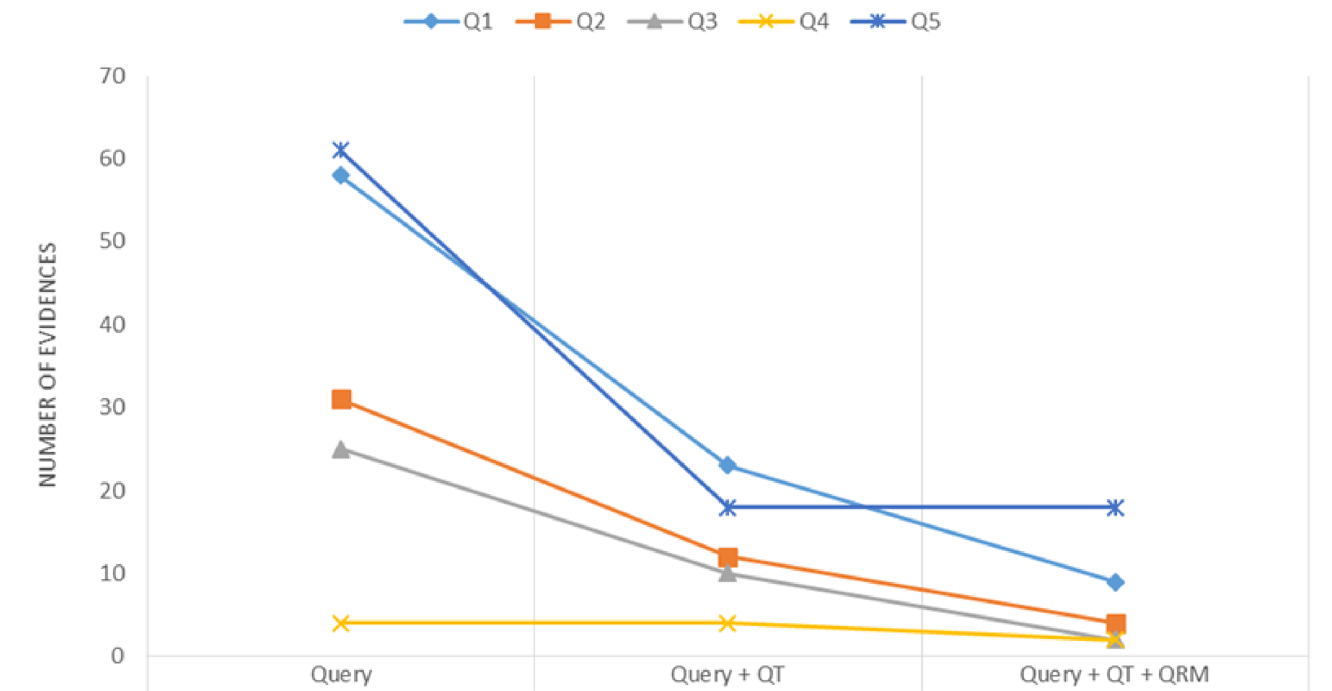


H = High	→ Highly Beneficial
M = Moderate	→ Moderate Beneficial
L = Low	→ Less Beneficial
U = Unknown	→ Unknown

Overall System Evaluation

Result evaluation for record reduction

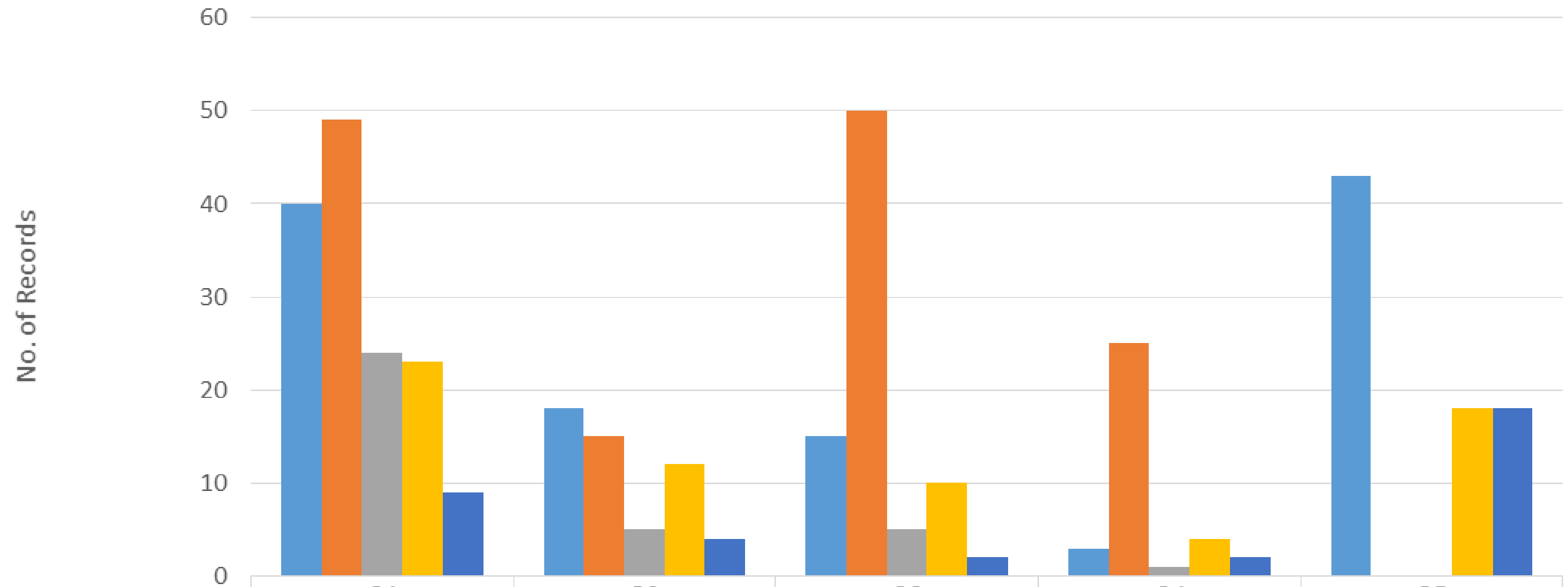
- On average, 51% records are reduced when clinical task (CT) is applied.
- Further, 48% records are eliminated on the average when QRM is applied.
- Overall, 75% records (on the average) are filtered out from the original query by applying CT and QRM.



	Query	Query + QT	Query + QT + QRM
Q1	58	23	9
Q2	31	12	4
Q3	25	10	2
Q4	4	4	2
Q5	61	18	18

Overall System Evaluation

Comparison with PubMed Derivative Systems



	Q1	Q2	Q3	Q4	Q5
RefMed	40	18	15	3	43
askMEDLINE	49	15	50	25	0
iPubMed	24	5	5	1	0
Clinical Queries	23	12	10	4	18
Proposed System	9	4	2	2	18

- Proposed system returned more accurate results for Q1, Q2, Q 3.
- Stands second for Q4
- Stands third for Q5, however, there were no results given by iPubMed and askMEDLINE.

Uniqueness and Contribution

Uniqueness
Contributions

Relevant Evidence

- Clinical Task Aware PICO Compliant Question Preparation with mean precision improved from 0.09 to 0.28 (about 3 times) and Mean Reciprocal Rank improved from 0.39 to 0.93 (about 2.5 times).
- Preprocessing, string matching, phrase-operator concatenation, and MeSH expansion

Quality Evidence

- Corpus preparation with no manual efforts for Quality Recognition Model
- Achieved 80.85% accuracy with standardized publication type feature which has improved the QRM accuracy by about 24%.

Contextually Fit Evidence

- Context Aware Grading (CAG) graded about 60% evidences as “High”.
- Achieved an agreement value of 0.37 (with human) which is fair enough for the experimental results.

Conclusion and Future Work

- Conclusion
  - Patient Data and Domain Knowledge/experience alone are not enough always for completing clinical decision process.
  - For improved and confident decision, it is required to acquire not only relevant rather quality evidences.
  - We proposed and experimented a methodology that supports methods of automatic evidence acquisition with PICO compliant question preparation and Grade the evidence on the basis of user context.
- Future Work
  - The work will progress to experiment the information extraction from the graded evidences for rule mining.
  - The algorithms developed for the accomplishment of this thesis can be extended to acquire "precision medicine" data.

Publications

- Patents (3)
  - Korean – 2 Published
  - International – 1 Applied
- SCI/E Journals (14)
  - First Author – 2 Published, 1 Minor Revision, 1 Major Revision
  - Co-Author – 9 Published, 1 Major Revision
- Non-SCI Journals (1)
  - Co-Author – 1 Published
- Conference (27)
  - First Author (10)
    - International (7)
    - Domestic (3)
  - Co-author (17)

Total Publications = 45

References

- {Sackett1996] Sackett, David L., et al. "Evidence based medicine: what it is and what it isn't."*BMJ: British Medical Journal* 312.7023 (1996): 71.
- [Aphinyanaphongs2005] Y. Aphinyanaphongs, I. Tsamardinos, A. Statnikov, D. Hardin, C.F. Aliferis, Text categorization models for high-quality article retrieval in internal medicine, J. Am. Med. Informatics Assoc. 12 (2005) 207–216. doi:10.1197/jamia.M1641.
- [Kilicoglu2009] H. Kilicoglu, D. Demner-Fushman, T.C. Rindflesch, N.L. Wilczynski, R.B. Haynes, Towards automatic recognition of scientifically rigorous clinical research evidence., J. Am. Med. Inform. Assoc. 16 (2009) 25–31. doi:10.1197/jamia.M2996.
- [Wilczynski 2005] Wilczynski NL, Morgan D, Haynes RB. An overview of the design and methods for retrieving high-quality studies for clinical care. BMC medical informatics and decision making l2005;5: 20.
- [Ebell2004] Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, Bowman M. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. The Journal of the American Board of Family Practice l2004;17: 59-67.
- [Grade2004] GRADE Working Group. "Grading quality of evidence and strength of recommendations." *BMJ: British Medical Journal* 328.7454 (2004): 1490.
- [Sarker2015] Sarker A, Moll D, Paris C. Automatic evidence quality prediction to support evidence-based decision making. Artificial intelligence in medicine l2015.
- [EBLIP2016] EBLIP, http://library.usask.ca/ceblip/eblip/the-steps-of-eblip.php, accessed on Jan 27, 2016.
- [EBBP2016] EBBP, http://www.ebbp.org/steps.html, accessed on Jan 27, 2016.
- [Verbert2012] Verbert K, Manouselis N, Ochoa X, Wolpers M, Drachsler H, Bosnic I, Duval E. Context-aware recommender systems for learning: a survey and future challenges. Learning Technologies, IEEE Transactions on l2012;5: 318-335.
- [Spark1976] Sparck Jones K, van Rijsbergen CJ. Information retrieval test collections. Journal of documentation l1976;32: 59-75.
- [Carletta1996] Carletta J. Assessing agreement on classification tasks: the kappa statistic. Computational linguistics l1996;22: 249-254.
- [Fontelo2005] Fontelo, Paul, Fang Liu, and Michael Ackerman. "askMEDLINE: a free-text, natural language query tool for MEDLINE/PubMed." *BMC medical informatics and decision making* 5.1 (2005): 5.
- [Rycroft2004] Rycroft-Malone, Jo. "The PARIHS Framework—A Framework for Guiding the Implementation of Evidence-based Practice." *Journal of nursing care quality*19.4 (2004): 297-304.
- [Boudin2010] Boudin, Florian, Lixin Shi, and Jian-Yun Nie. "Improving medical information retrieval with PICO element detection." *European Conference on Information Retrieval*. Springer Berlin Heidelberg, 2010.
- [Young 2009] Young, Jane M., and Michael J. Solomon. "How to critically appraise an article." Nature Clinical Practice Gastroenterology & Hepatology 6.2 (2009): 82-91.

Thank You!

Any Question or Comments?