IBM

Smarter education for smarter planet

IBM Innovation Centre for Education

Emerging Applications of Natural Language Generation in Information Visualization, Education, and Health Care

Unit objectives

After completing this unit, you should be able to:

- Gain knowledge on the process of Multimedia Presentation Generation
- Learn the concept of Language Interfaces for Intelligent Tutoring Systems
- Gain an insight into Argumentation for Healthcare Consumers
- Learn the concepts of Clinical Decision Support Systems
- Understand the core concepts of Sentiment Analysis and Subjectivity

Multimedia presentation generation

- Multimedia generation ➔ Combining the various forms of media.
- Images, audio and video in the results.
- Corpora ➔ Rich Layout, diagrams.
- Generators ➔ Enhanced ➔ Multimedia content.

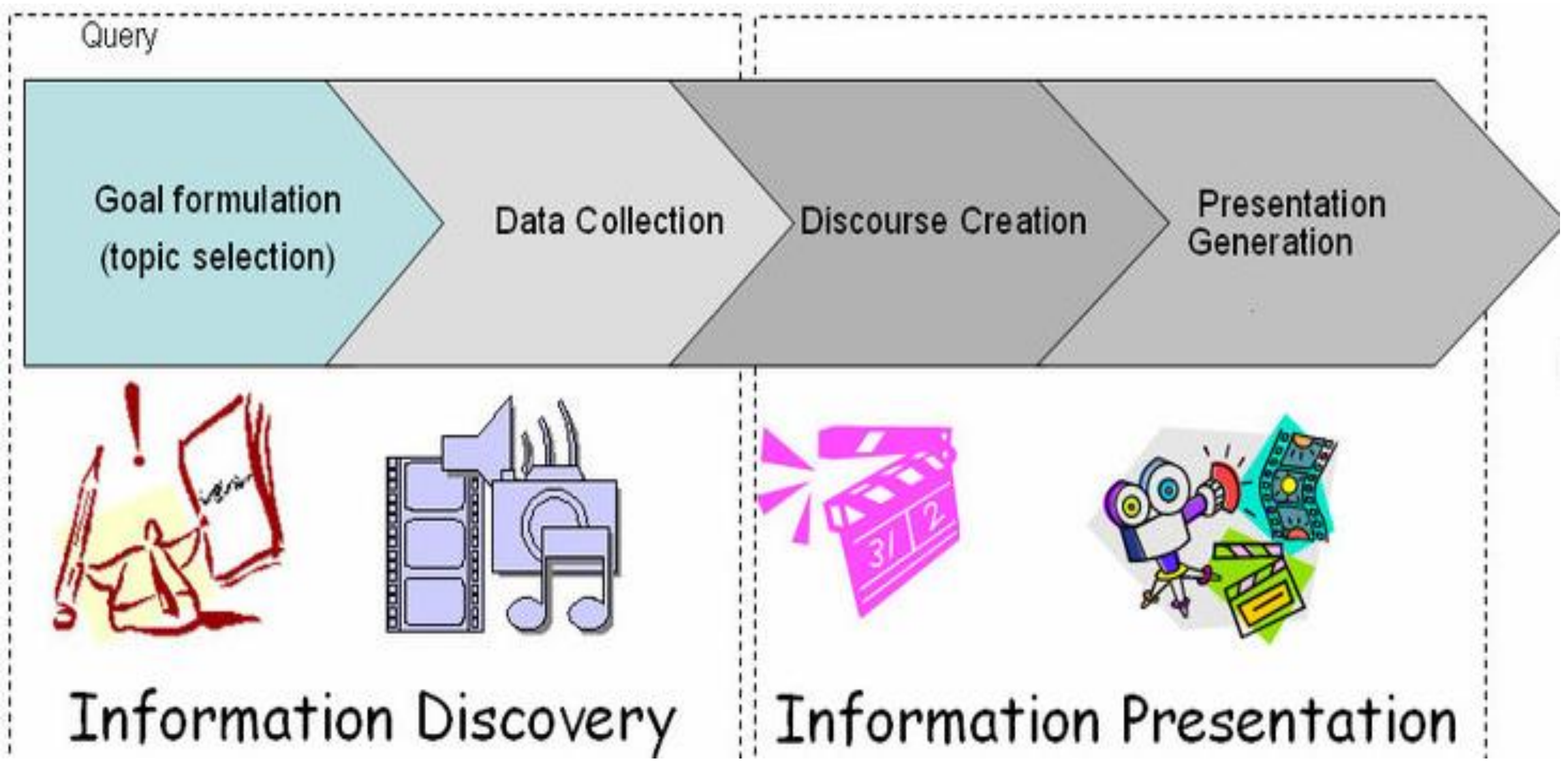


Figure: Multimedia Presentation Generation Outline
Source: https://www.researchgate.net/figure/Presentation-generation-steps-in-MANA_fig2_259922670

Focus points to add multimedia in NLG

- Meaning of the text.
- Style.
- Wording of the text.

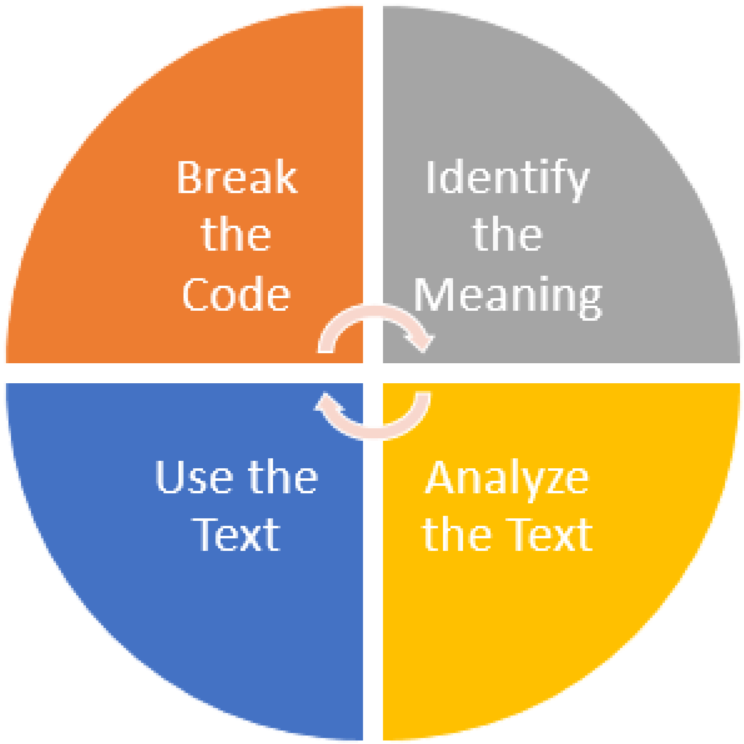


Figure: Text Meaning

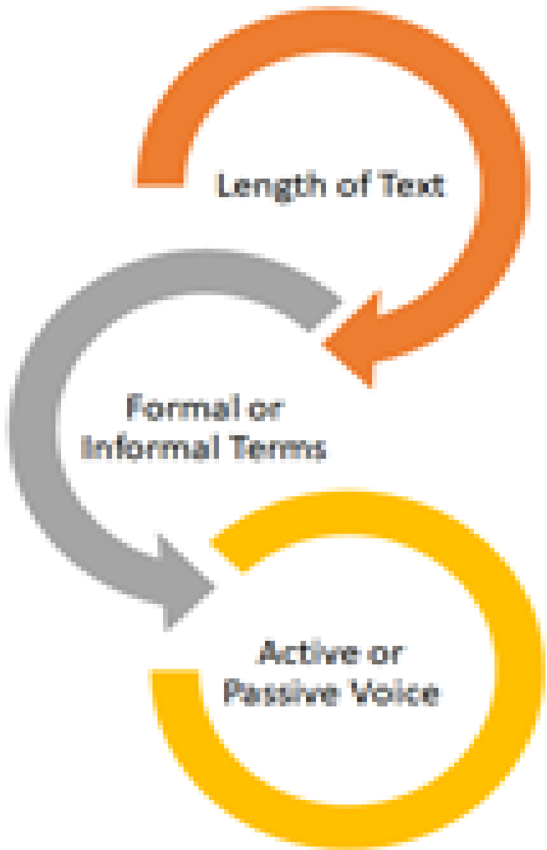
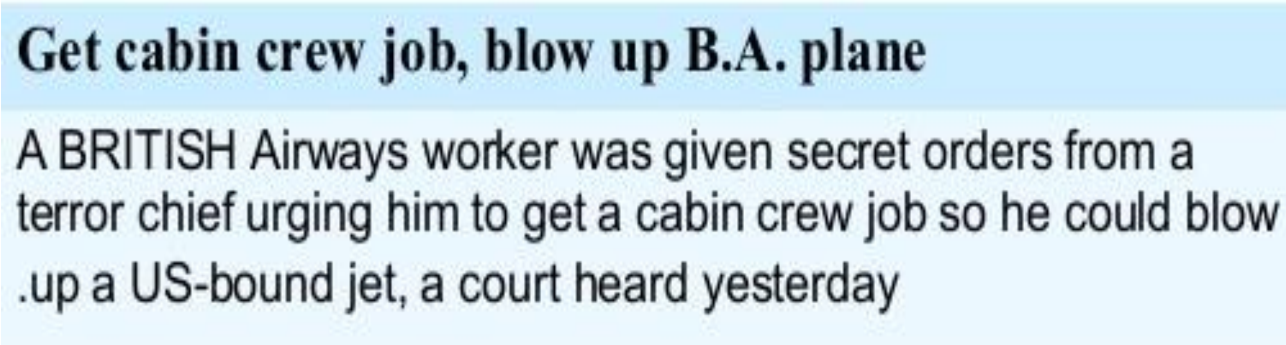


Figure: Text Style



Get cabin crew job, blow up B.A. plane

A BRITISH Airways worker was given secret orders from a terror chief urging him to get a cabin crew job so he could blow .up a US-bound jet, a court heard yesterday

Figure: Usage of Words

Text generation: Meaning representation

- Sample representation:
  - obligatory(s:suggest(y:doctor(z:patient),d:dose))
  - obligatory(follow(z,s))
  - [unsure-about(z,d) v unsure-about(z,timing(d))]-> obligatory(ask(z,y))
  - procedure(take(z,t:tablet), [remove(z,t,foil,finger,back(t))& swallow(z,t,water)])
  - [w:take(z,overdose)] -> obligatory([tell(z,y,w) OR visit(z,casualty(hospital(z)))])
  - store(a:person,m:medicine) -> obligatory(storeawayfrom(a,m,children))

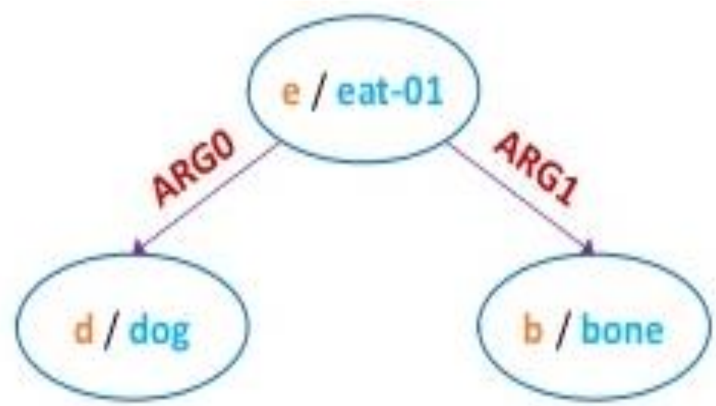
- Nodes are variables labelled by concepts
  - Entities, events, states, properties
  - d / dog: d is an instance of dog
- Edges are semantic relations

Figure: Text representation

"The dog is eating bones."

```
(e / eat-01
    :ARG0 (d / dog)
    :ARG1 (b / bone))
```

Figure: Parameterized representation



eat.01: *consume* (VN-class: eat-39.1, FN-frame: Ingestion)

ARG0-PAG: *consumer*, *eater* (VN-role: agent)

ARG1-PPT: *meal* (VN-role: patient)

Figure: Structured representation

Text generation: Document structure design (1 of 2)

- Natural language generation systems ➔ Semantic rules ➔ Aggregation.
- Left to Right pattern.
- Complex sentences ➔ Aggregation.
- Structure ➔ Clauses, Sentences and Paragraphs.

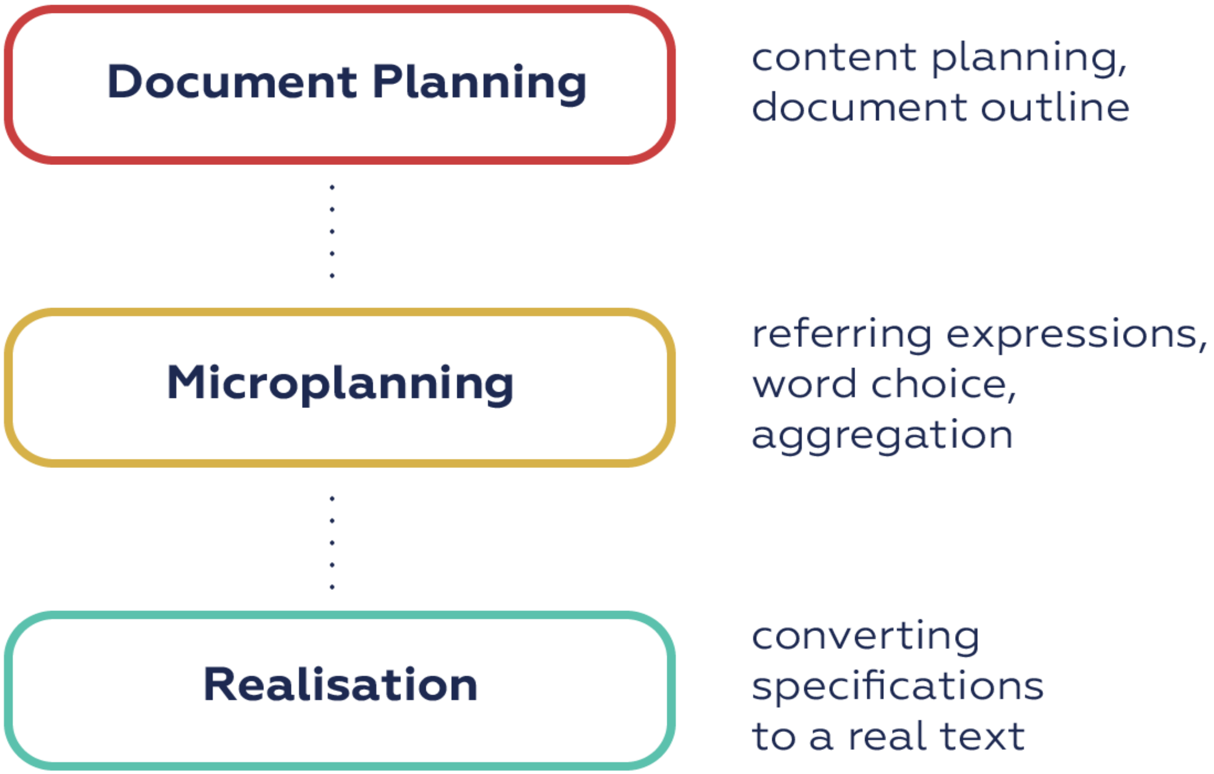


Figure: Text generation: Document structure design
Source: https://medium.com/sciforce/a-comprehensive-guide-to-natural-language-generation-dd63a4b6e548

Text generation: Document structure design (2 of 2)

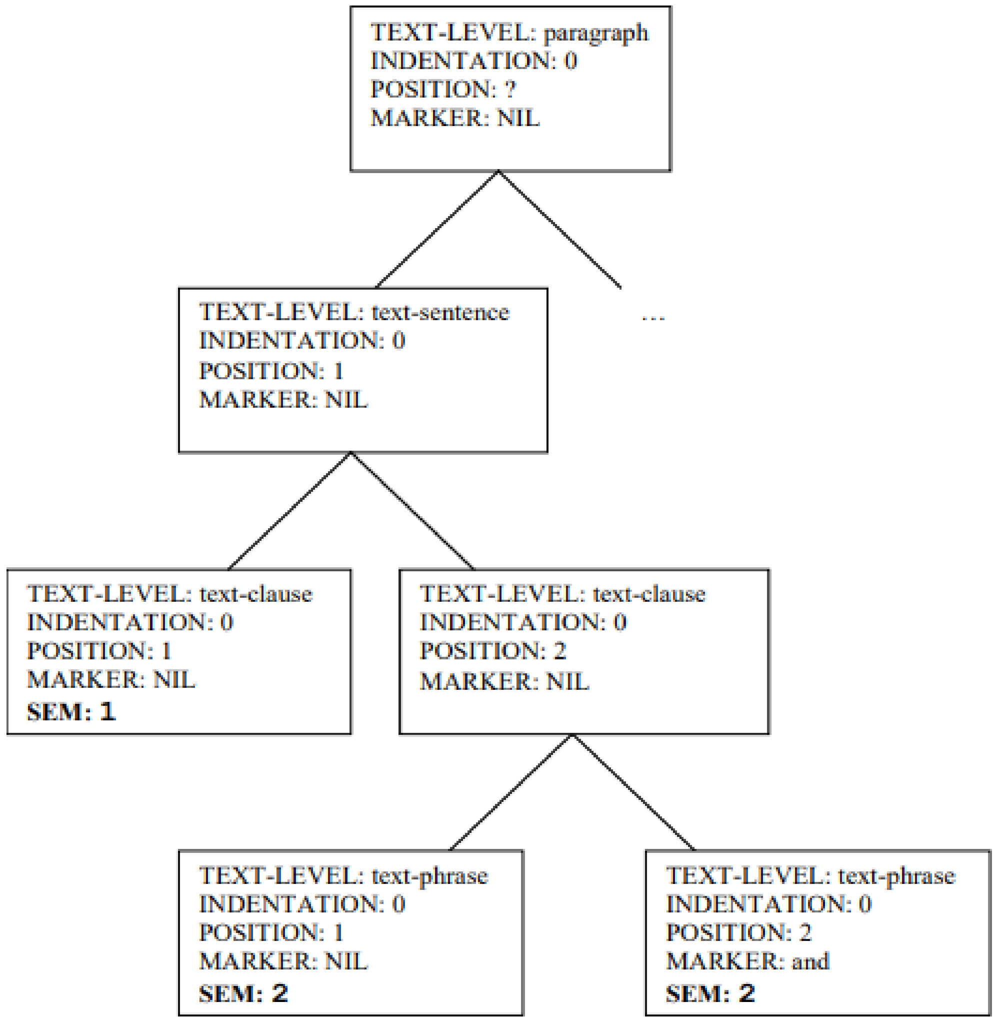


Figure: Document Representation

Source: https://www.researchgate.net/publication/227056814_Generating_Multimedia_Presentations_from_Plain_Text_to_Screen_Play

Text generation: Linguistic style control

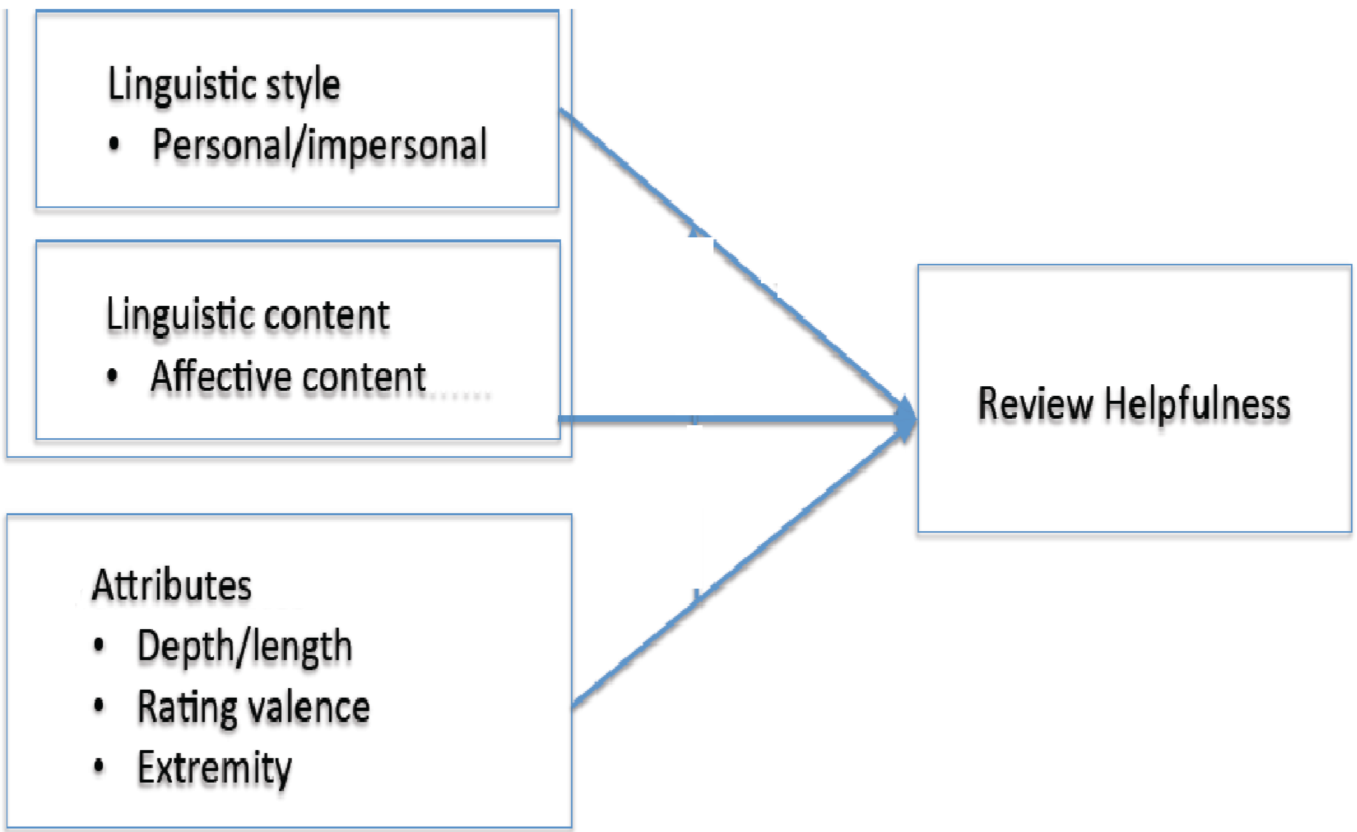


Figure: Linguistic Style Control

Source: https://www.semanticscholar.org/paper/Linguistic-Style-and-Online-Review-Helpfulness-Wang-Karimi/4bf204a94c6991c56f47f47f0a471c1f0d48d22c

Document layout

Figure: Document layout formats

Source: https://www.vectorstock.com/royalty-free-vector/document-report-layout-templates-set-vector-9437687

Layout and meaning representation

- Mapped to the abstract document structure.
- Simple words put together ➔ Meaning.
  - Uniformity.
  - Size.
  - Length.
  - Navigability.
  - Spacing.

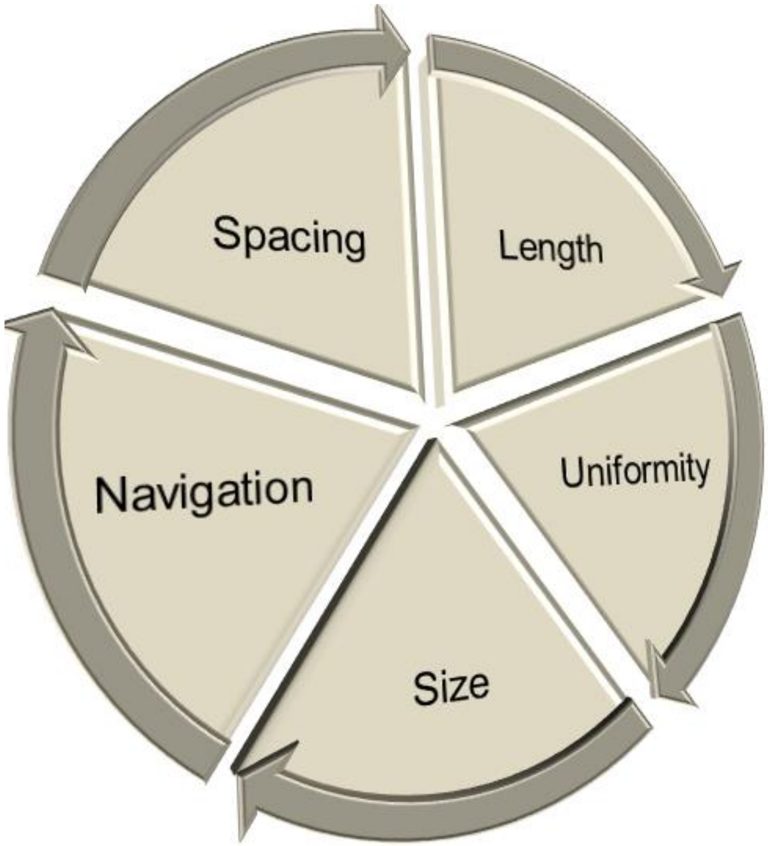


Figure: Layout and meaning factors

- Consult your doctor if you have any problems concerning your treatment, or any questions about your treatment.

Layout style and wording representation

- Natural language generation systems ➔ Formatting ➔ Syntax.
- Layout and Wording interact with one another ➔ Efficient communication.
- Change in Style ➔ Alter the words.

Sample Text:

Ask your doctor if:

• you are unsure about the dose

• you are unsure when to take the dose

Change in style without change in wording

Ask your doctor if: you are unsure about the dose you are unsure when to take the dose.

Change in style with change in wording

Ask your doctor if you are unsure about the dose or you are unsure when to take the dose.

Figure: Change in Style with Wording

Image style and meaning representation

- Meaning of a picture ➔ Relative and Subjective.
- Based on conjunction.
- Library of pictures.
- Representation of images ➔ Understanding of the context.
  - Complex piece of information.
  - Events that should be followed sequentially.
  - Continuous quantities.
  - Presentation of serial element.

x:person & y:medicine & getmedicine(x,y)

x:person & y:tablet & remove(x,y,foil,finger,back(y))

Figure: Picture as Representation

Figure: Image representation

Image and wording usage

- Images become an important part of the document.
- Enhance the power of the document.
- References to the illustrative images, Reduce the number of words.
- References ➔ Provide connectivity.

Take your tablet by removing it from the foil by pressing your against the back of the tablet

Take your tablet by removing it from the foil

Figure: Sample Representation

Scripted dialogue (1 of 2)

Documents ➔ Objects, Dialogues ➔ Events.

Narrative and Argumentative representation.

Message ➔ "m" Signal ➔ "s".

Plain text representation

Although the patient asked when he should take the medicine, the pharmacist could only reply to him that his doctor would be able to tell him.

Dialogue representation

Patient: When should I take the medicine?

Pharmacist: Your doctor will be able to tell you.

Figure: Dialogue based representation

Scripted dialogue (2 of 2)

- Dialogue planner:
  - Multimodal generator.
  - Speech synthesis.
  - Gesture Assignment.
  - Media player.

Pharmacist:	Here is your medicine.
	Store it away from children.
	Your doctor should suggest a dose.
Patient:	Can I change the dose?
Pharmacist:	No, you should follow your doctor's advice.
Patient:	When should I take the medicine?
Pharmacist:	Ask your doctor.

Figure: Dialogue instead of plain text

Language interfaces for intelligent tutoring systems

IBM ICE (Innovation Centre for Education)

- Intelligent tutoring systems ➔ Lessons without intervention from human.
- Goal ➔ Learn and Teach.
  - User interface.
  - Pedagogical model.
  - Domain knowledge model.
  - Student model.

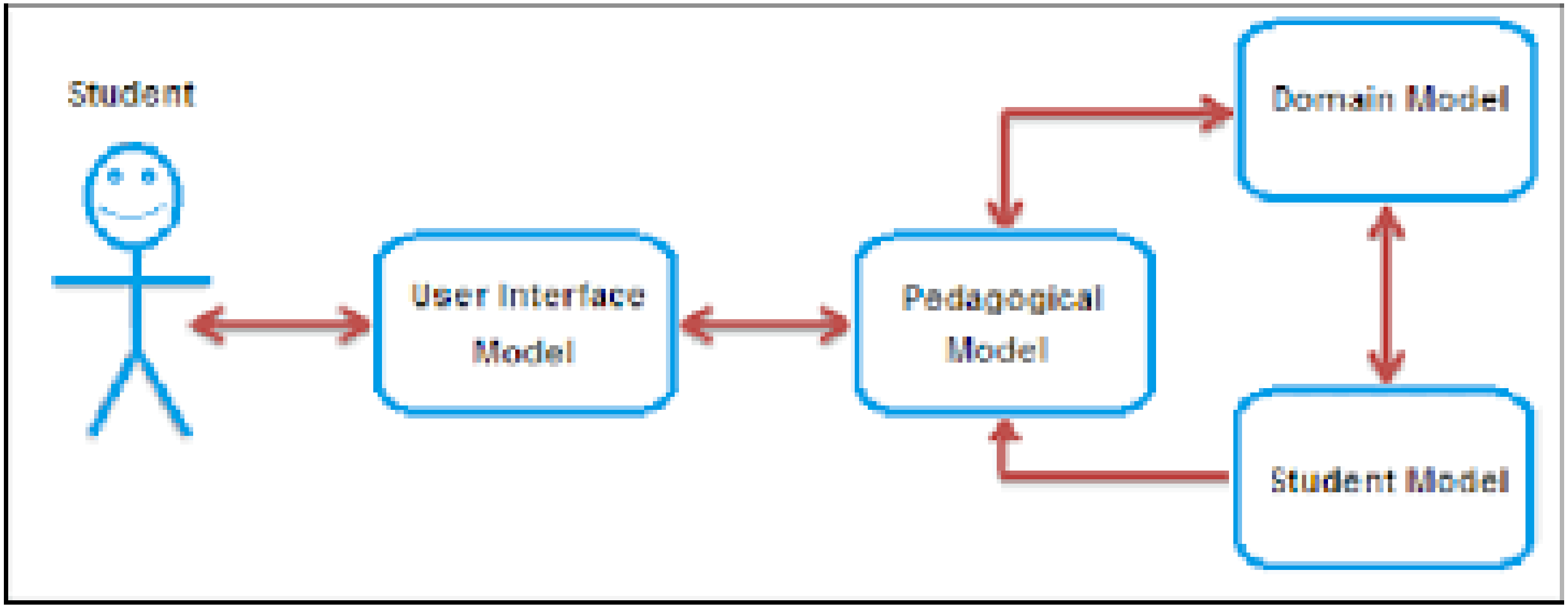


Figure: A simple Tutoring system

Source: https://www.researchgate.net/figure/Typical-architecture-of-the-Intelligent-Tutoring-System_fig1_314229691

CIRCSIM-Tutor

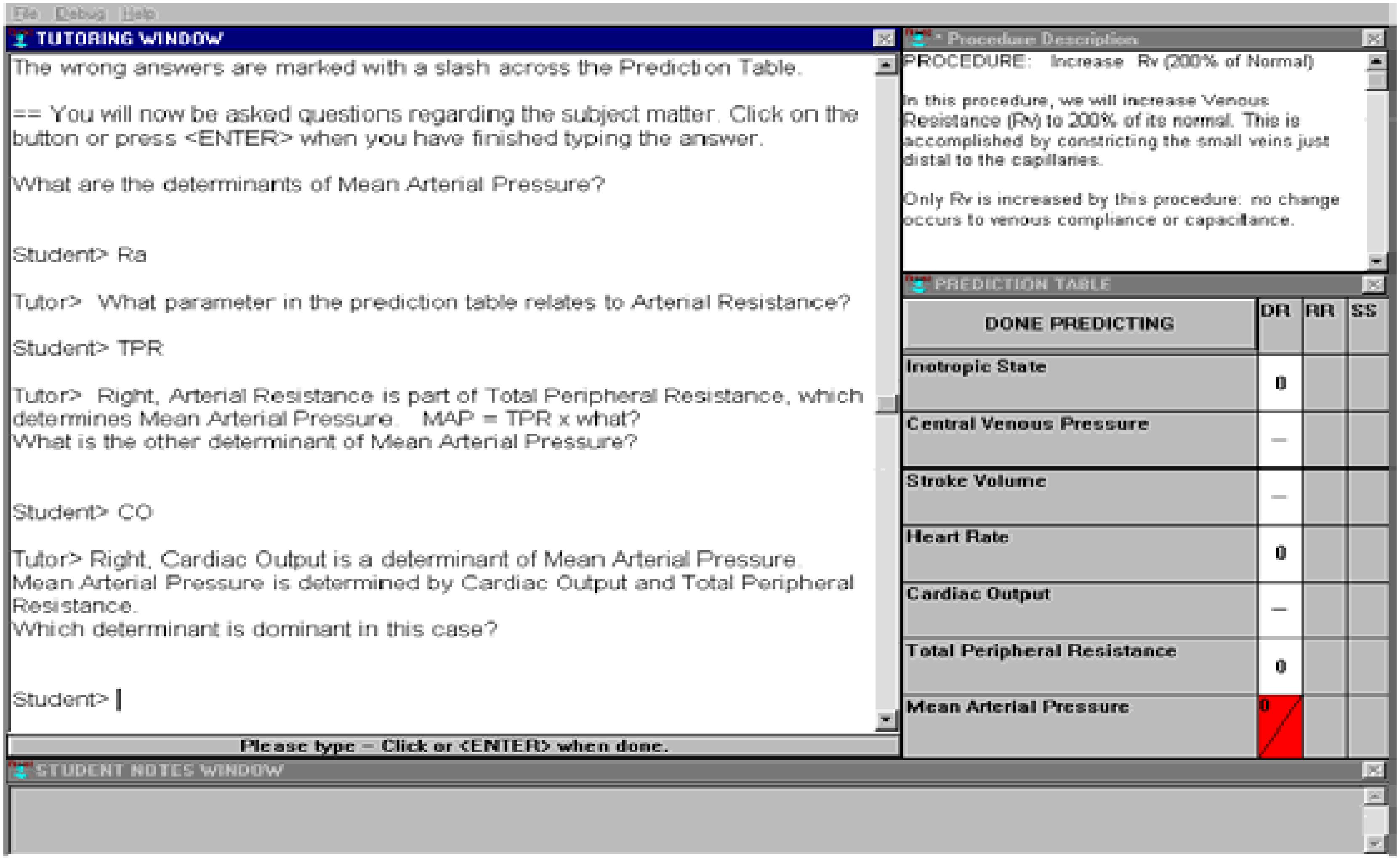


Figure: Screen Interface of CIRCSIM-Tutor

Source: https://www.researchgate.net/figure/CIRCSIM-Tutor-screen-containing-a-fragment-of-dialogue-from-Session-39-CIRCSIM-Tutor_fig1_228342817

CIRCSIM-Tutor architecture, data presentation and process cycle (1 of 2)

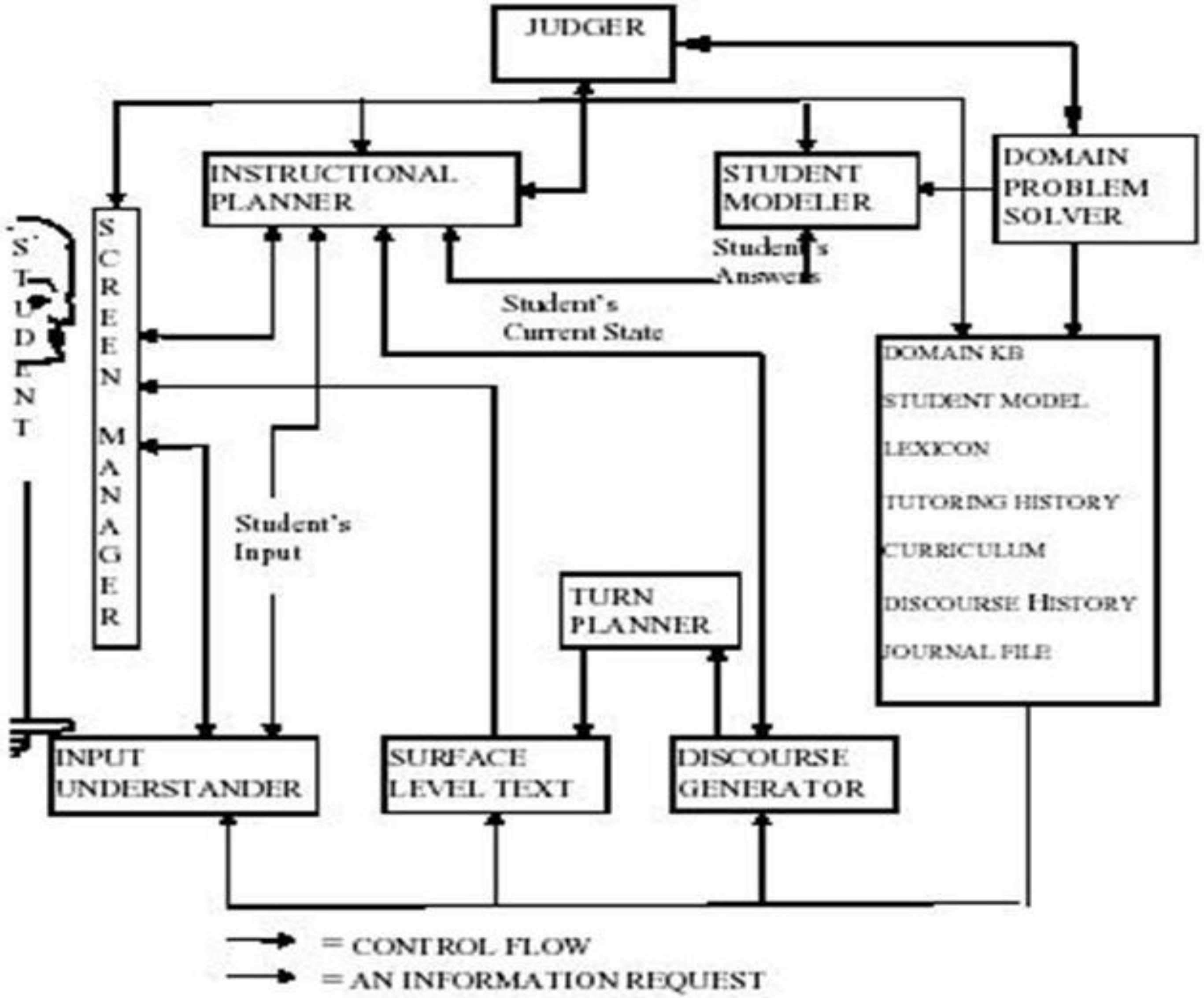


Figure: CIRCSIM-Tutor Architecture

Source: https://slideplayer.com/slide/3082681/

CIRCSIM-Tutor architecture, data presentation and process cycle (2 of 2)

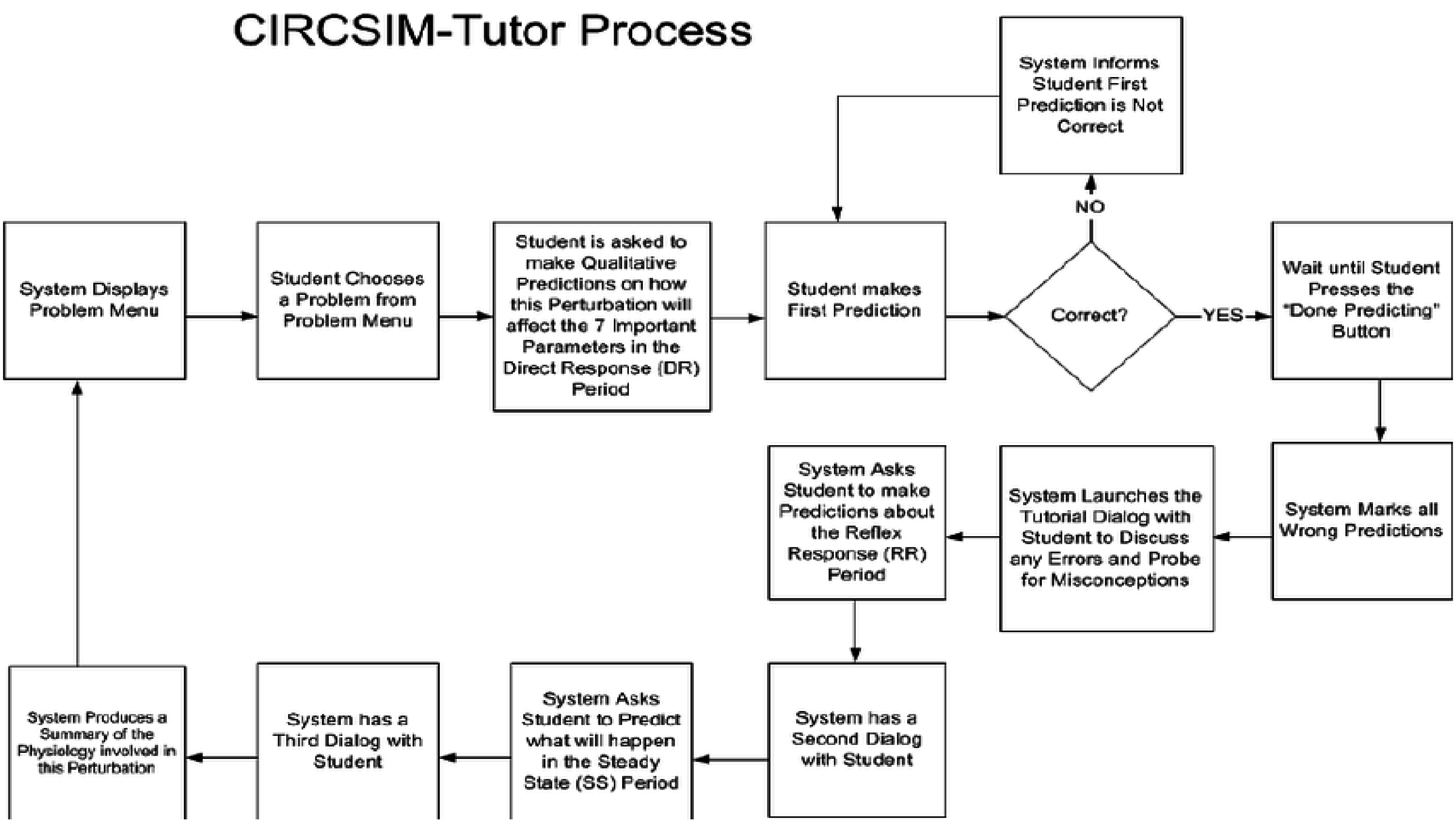


Figure: Process steps

Source: https://www.researchgate.net/figure/The-Tutoring-Process-in-CIRCSIM-Tutor_fig4_228797088

AUTOTUTOR

- Auto tutor ➔ Science, Technology.
- Animated conversational agent, dialogue management, electronic documents.
- Topmost area ➔ Problem.
- Left area ➔ Conversational agent.
- Right area ➔ Auxiliary diagrams, Student interactions.

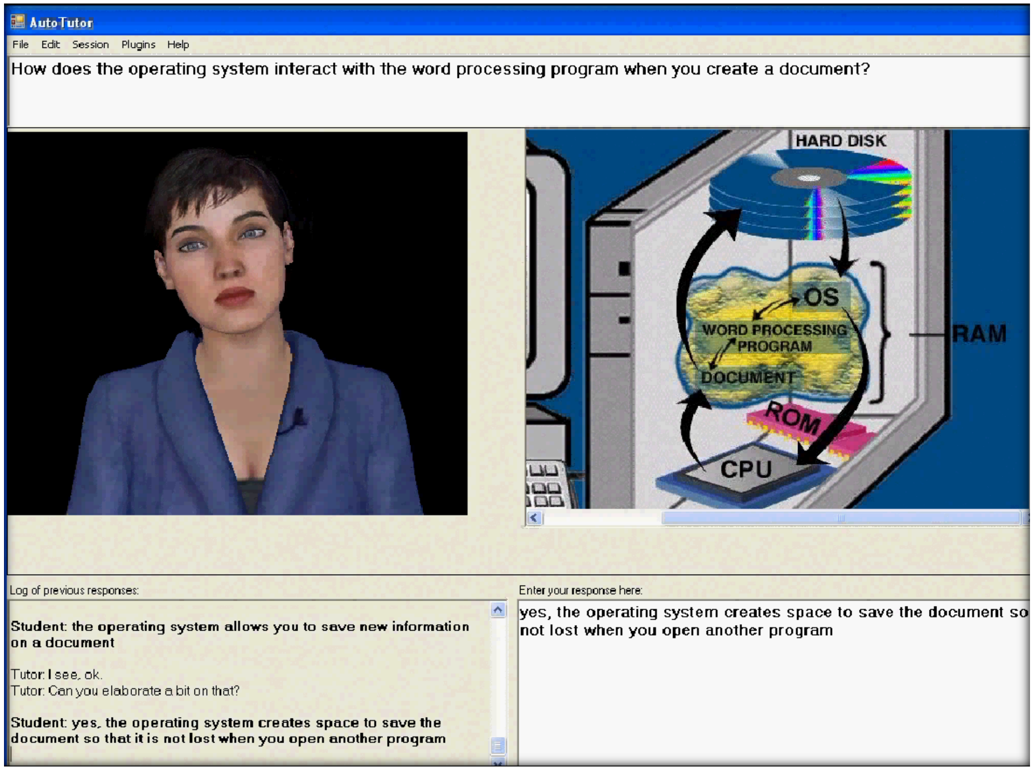


Figure: User interface

Source: https://www.semanticscholar.org/paper/AutoTutor-and-affective-autotutor%3A-Learning-by-with-D%27Mello-Graesser/a884f00fe18a6abf837b2ccb490165ded90fc29a

AUTOTUTOR architecture and process

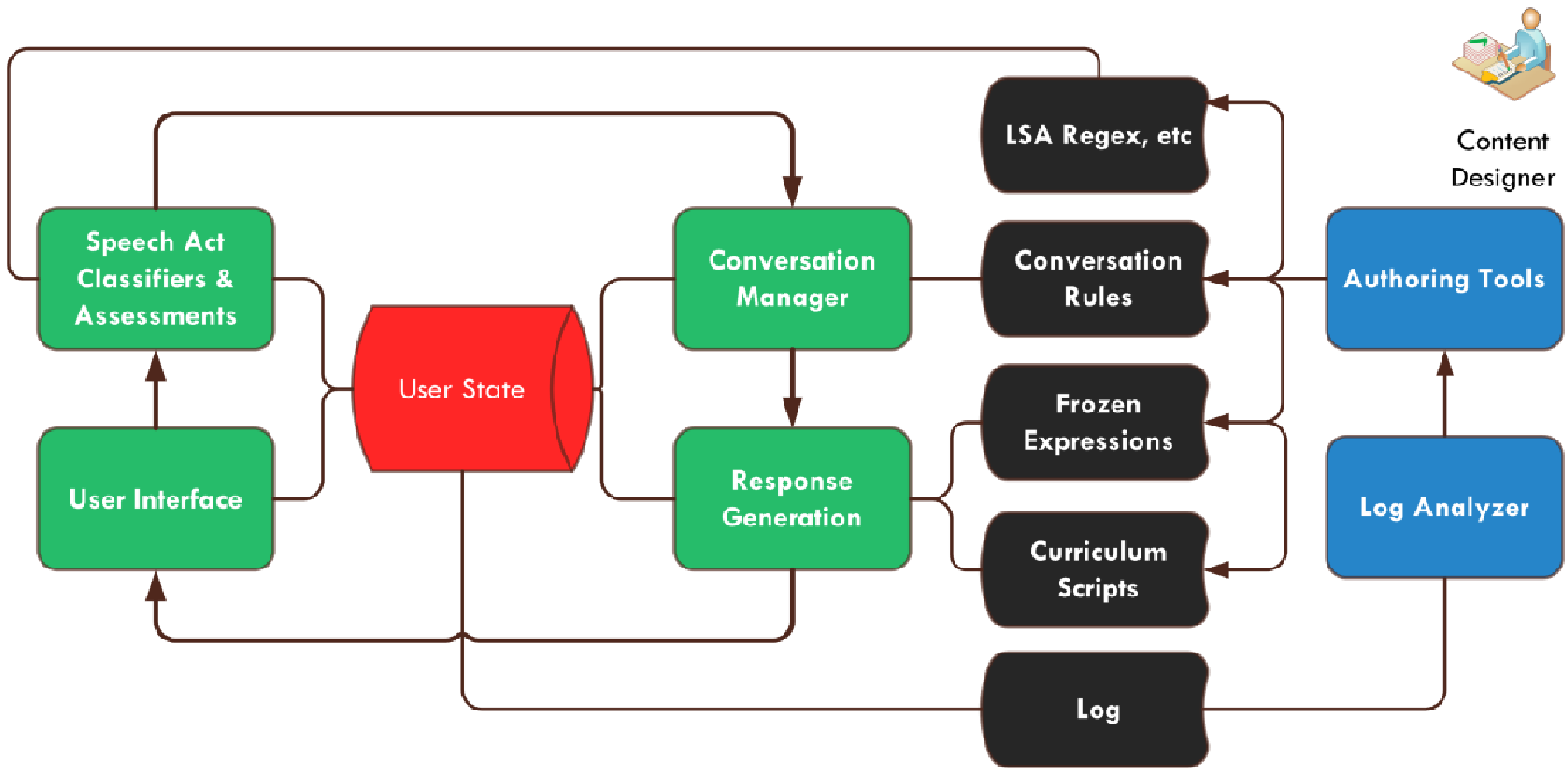


Figure: AUTOTUTOR architecture

Source: https://www.semanticscholar.org/paper/AutoTutor-and-affective-autotutor%3A-Learning-by-with-D'Mello-Graesser/a884f00fe18a6abf837b2ccb490165ded90fc29a/figure/1

ATLAS Andes

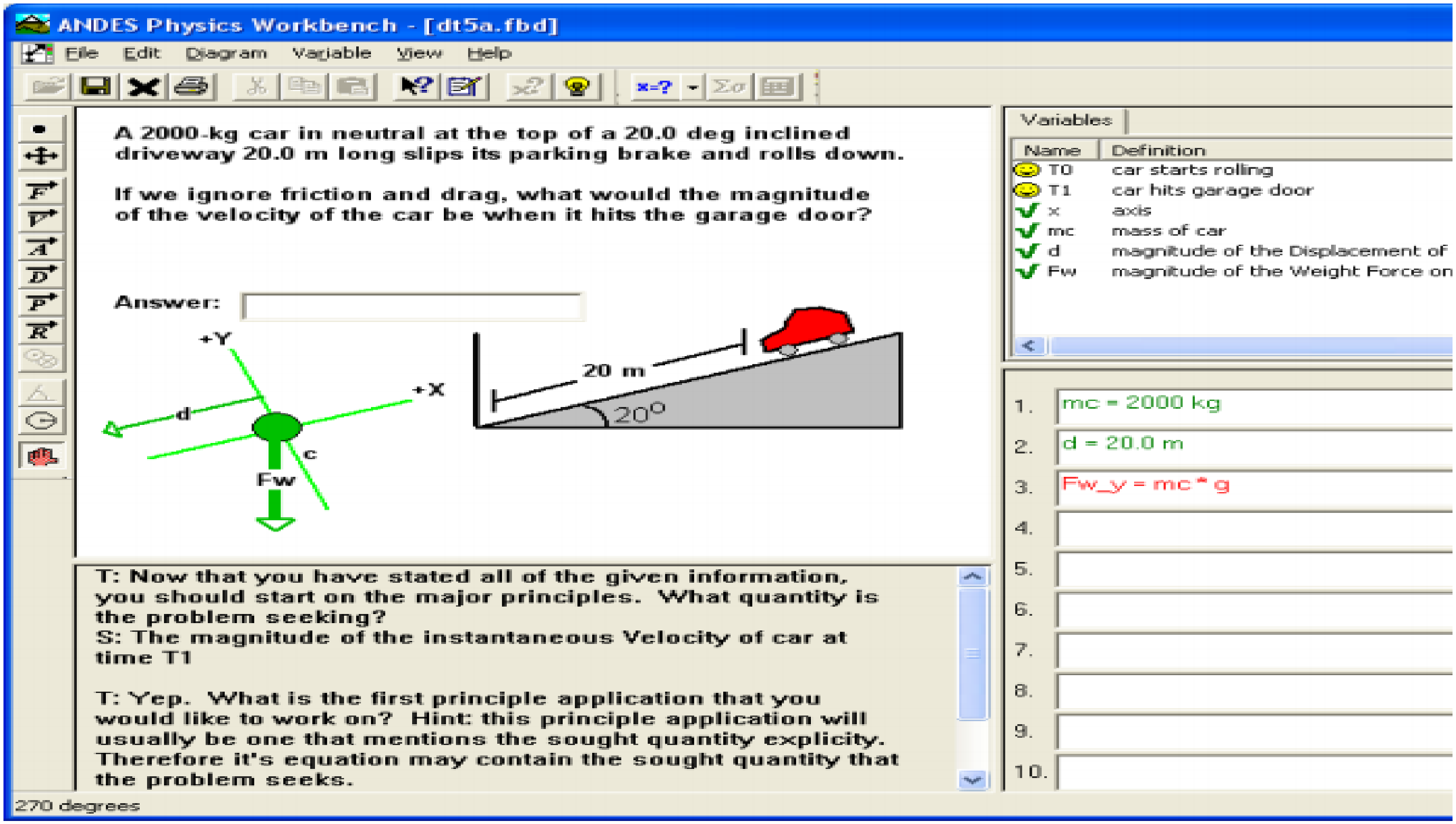


Figure: Andes User Interface

Source: http://www.andestutor.org/Pages/AndesLessonsLearnedForWeb.pdf

Andes system architecture and design (1 of 2)

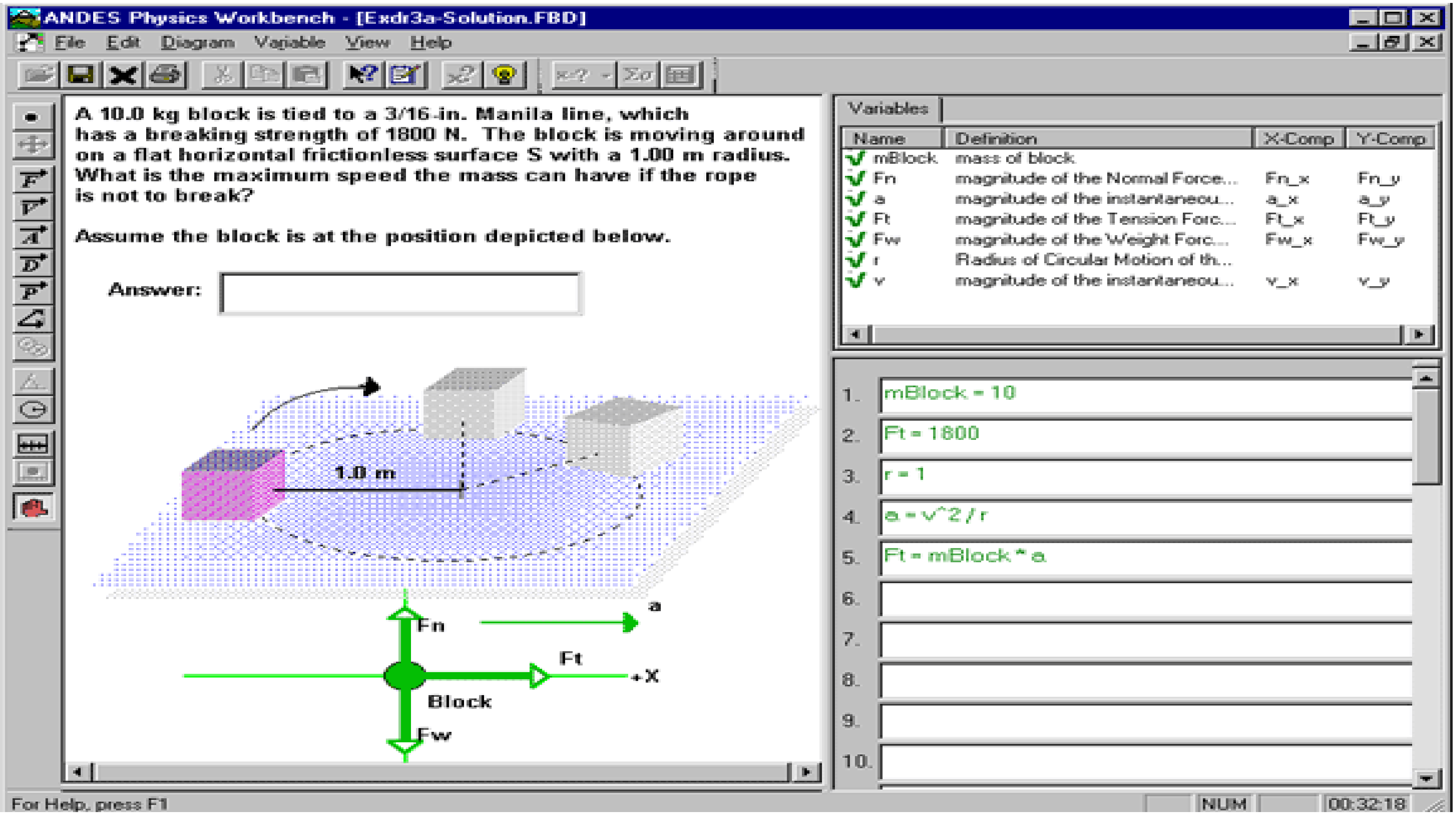


Figure: Work Environment

Source: https://quod.lib.umich.edu/j/jep/3336451.0006.110?view=text;rgn=main

Andes system architecture and design (2 of 2)

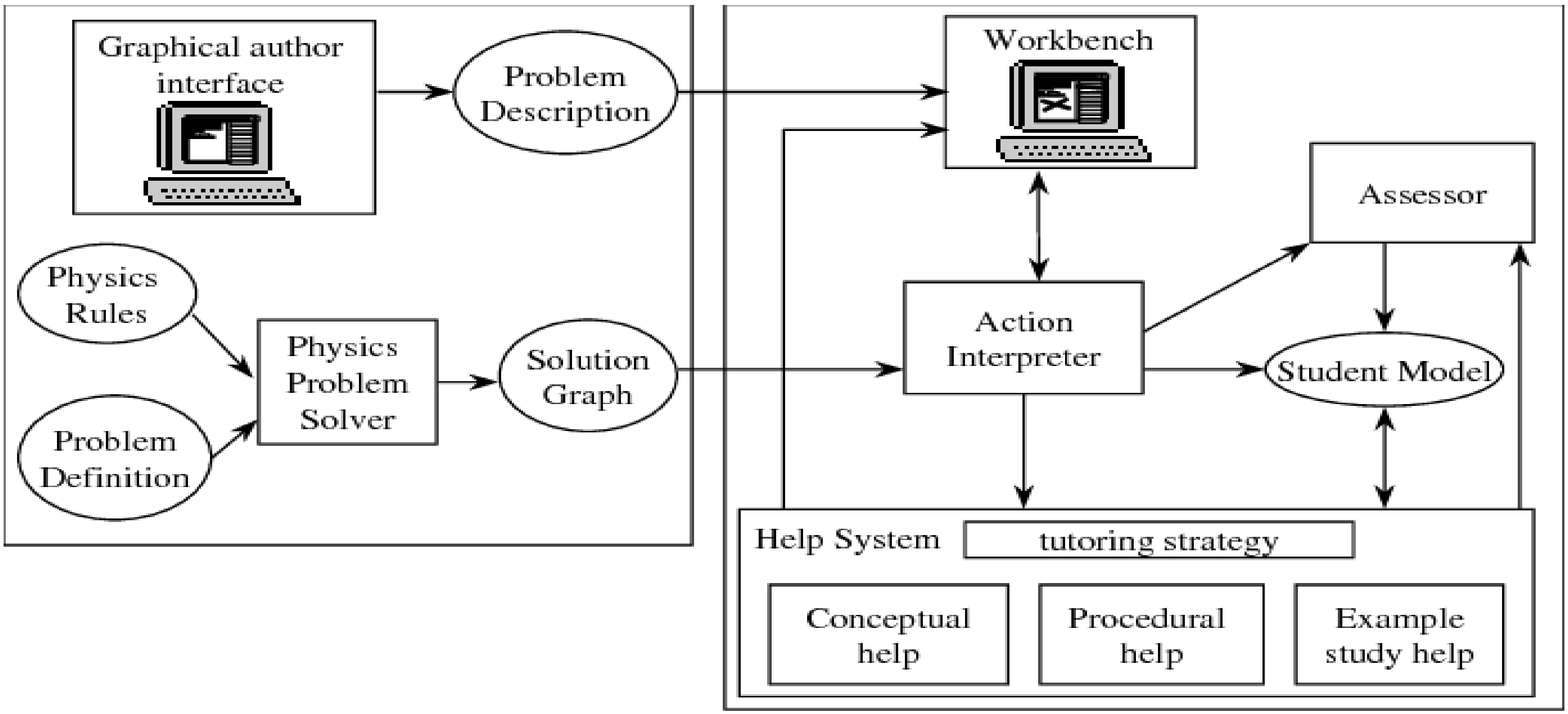


Figure: Architecture of Andes

Source: https://www.researchgate.net/figure/The-Andes-System-Architecture-Rectangles-are-system-modules-ellipses-are-data_fig1_2829113

-

Pedagogical considerations in Andes

IBM ICE (Innovation Centre for Education)

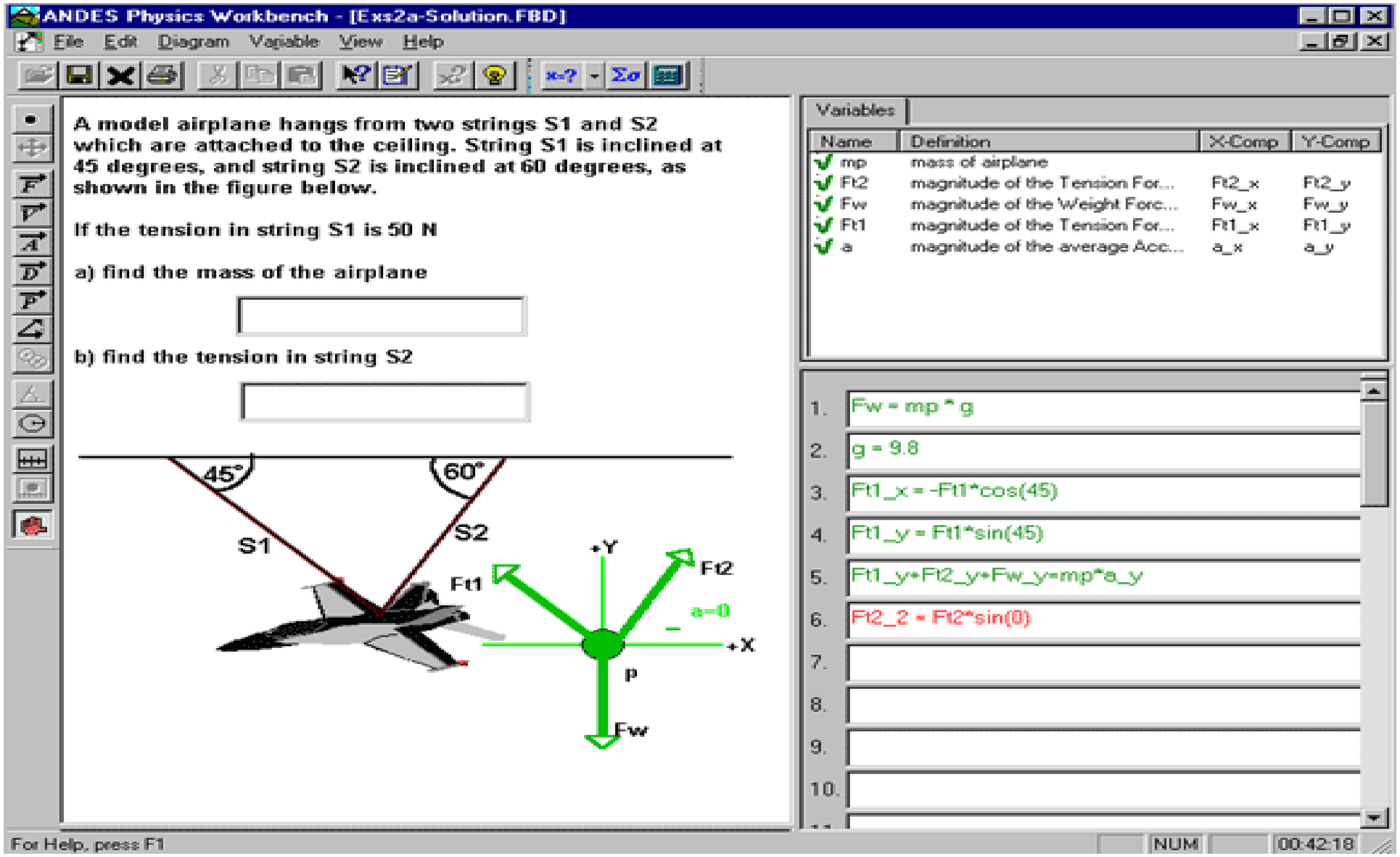


Figure: Multiple Solution Paths

Source: https://quod.lib.umich.edu/j/jep/3336451.0006.110?view=text;rgn=main

WHY2-ATLAS

Figure: Atlas User Interface

Source: http://i-publisher.atlasproject.eu/atlas/documentation/advanced/interface

Why 2 Atlas architecture and process

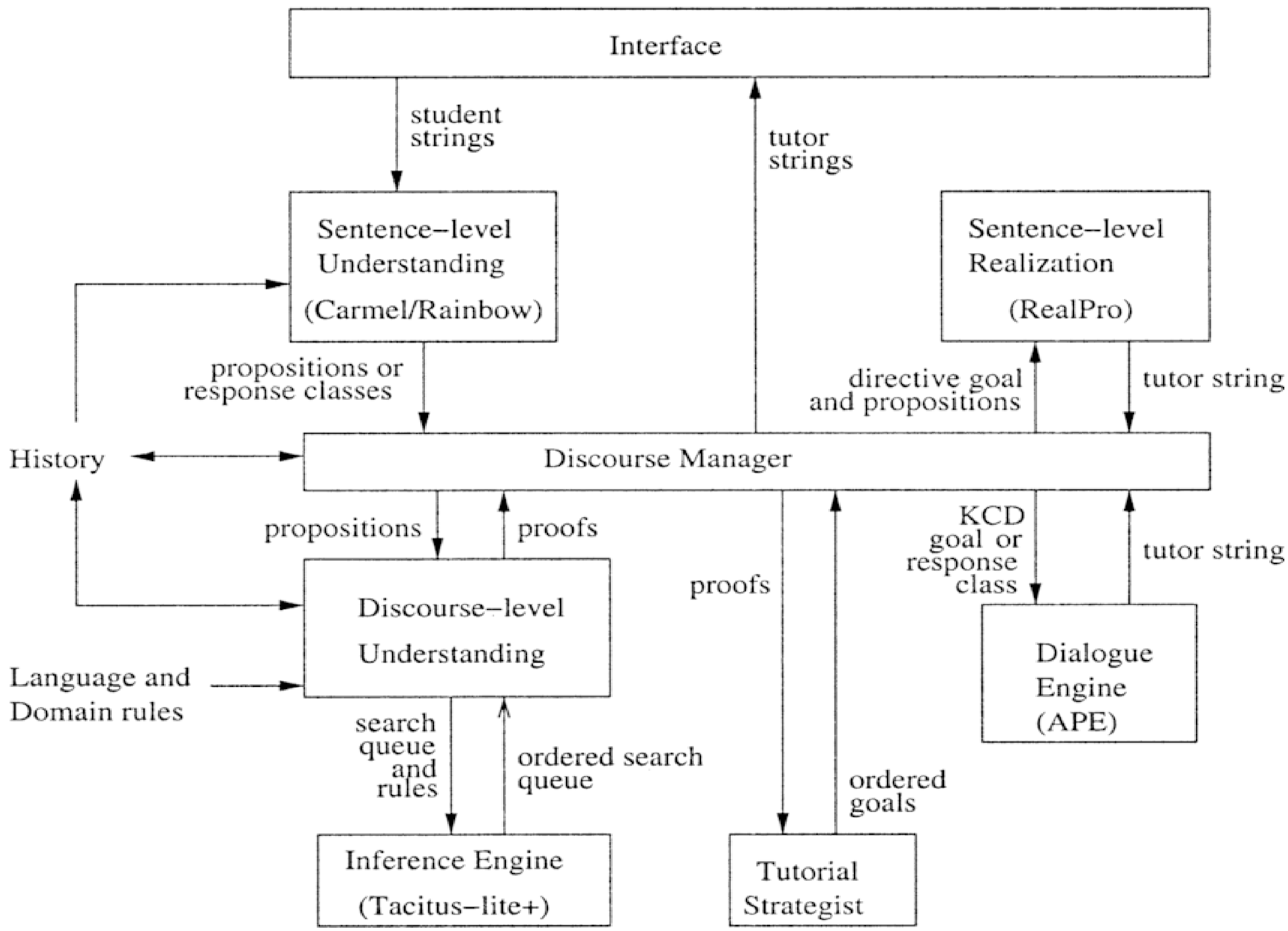


Figure: Architecture

Source: https://www.researchgate.net/figure/Why2-Atlas-tutoring-system-architecture_fig1_220532234

Argumentation for healthcare consumers (1 of 2)

IBM ICE (Innovation Centre for Education)

- Clinical Decision Support systems (CDS) ➔ Automating health related information.
- Goal ➔ Clinical decision with knowledge and accuracy.
- Patient data Integration Centre:
  - Decision rules.
  - Knowledge base.
  - Assess.
  - Generate recommendation.

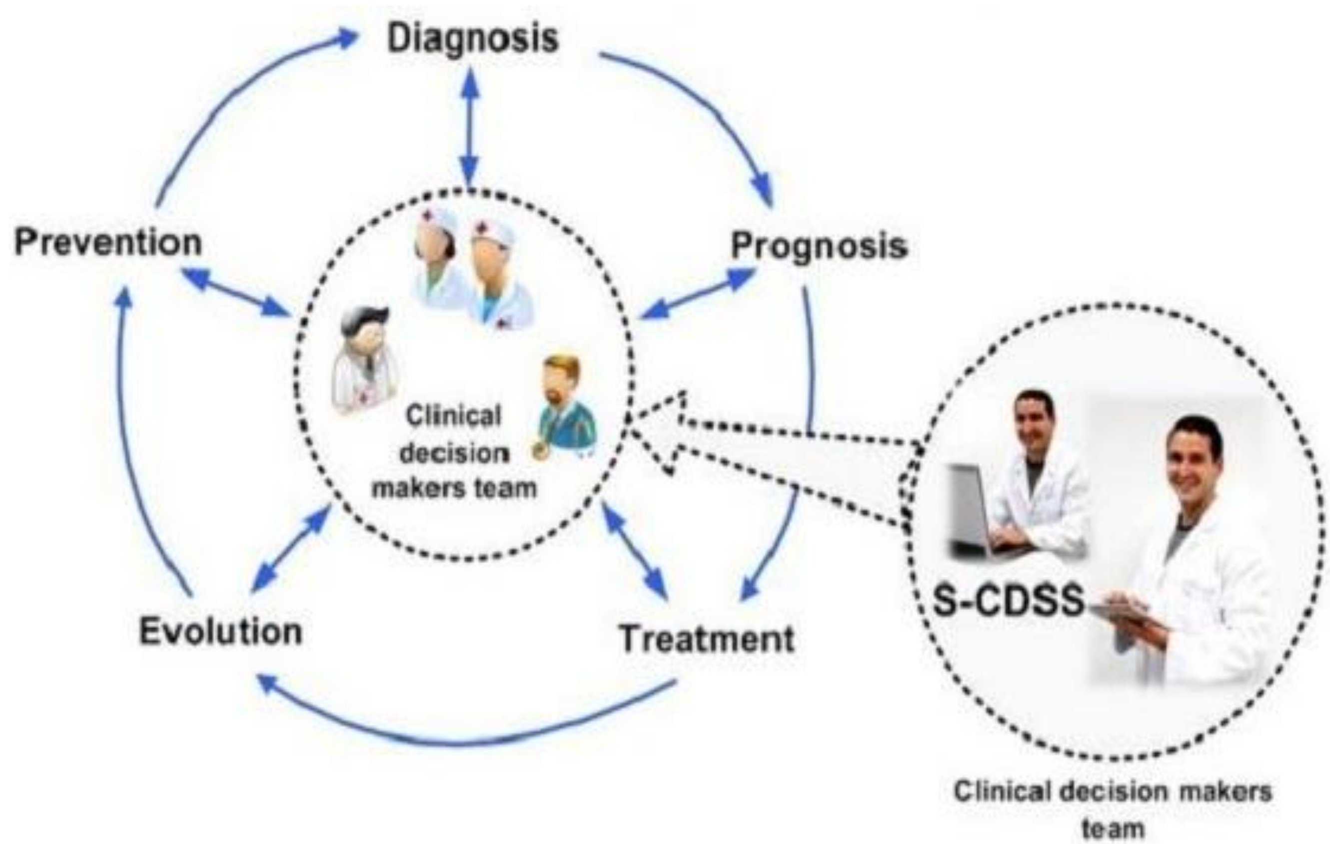


Figure: CDS Outline

Source: https://www.kenresearch.com/blog/2019/11/global-clinical-decision-support-systems-market/

Argumentation for healthcare consumers (2 of 2)

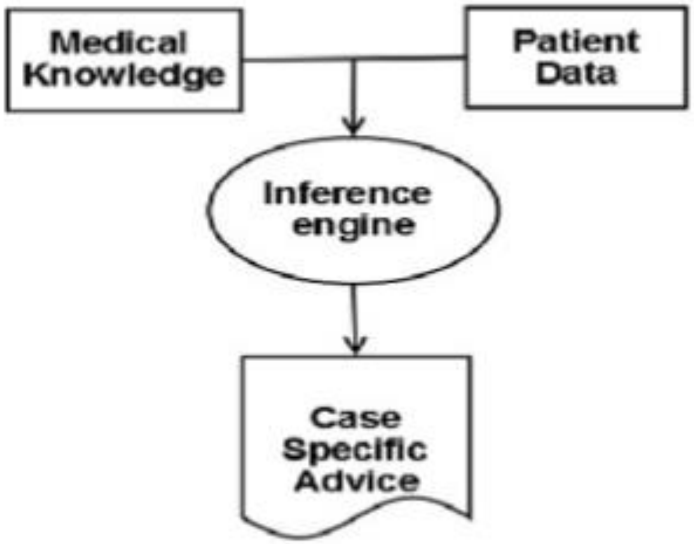


Figure: CDS Components

Source: https://www.medgadget.com/2018/07/clinical-decision-support-systems-cdss-market-prepare-to-touch-at-a-cagr-of-11-5-by-2023.html

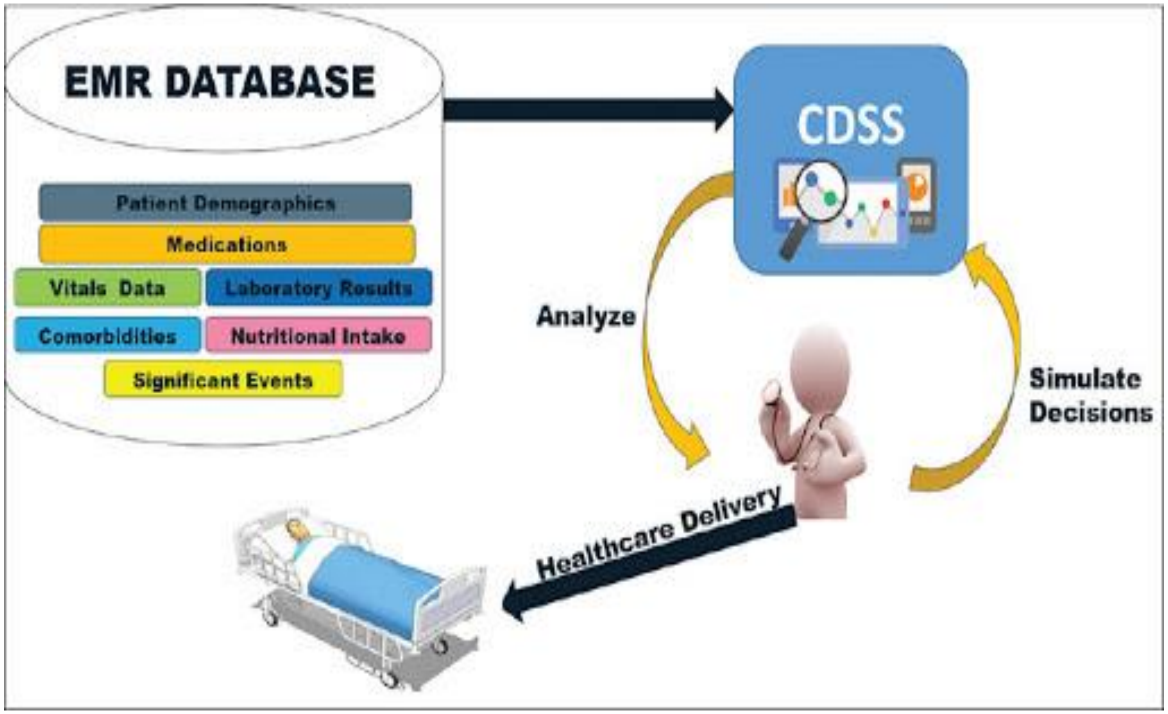


Figure: CDS – EHR/EMR Interaction

Source: http://www.ijam-web.org/article.asp?issn=2455-5568; year=2017; volume=3; issue=1; spage=78; epage=83;aulast=Pappada

CDS architecture and processing

- Major activities:
  - Alerting.
  - Monitoring.
  - Coding.
  - Reminders.

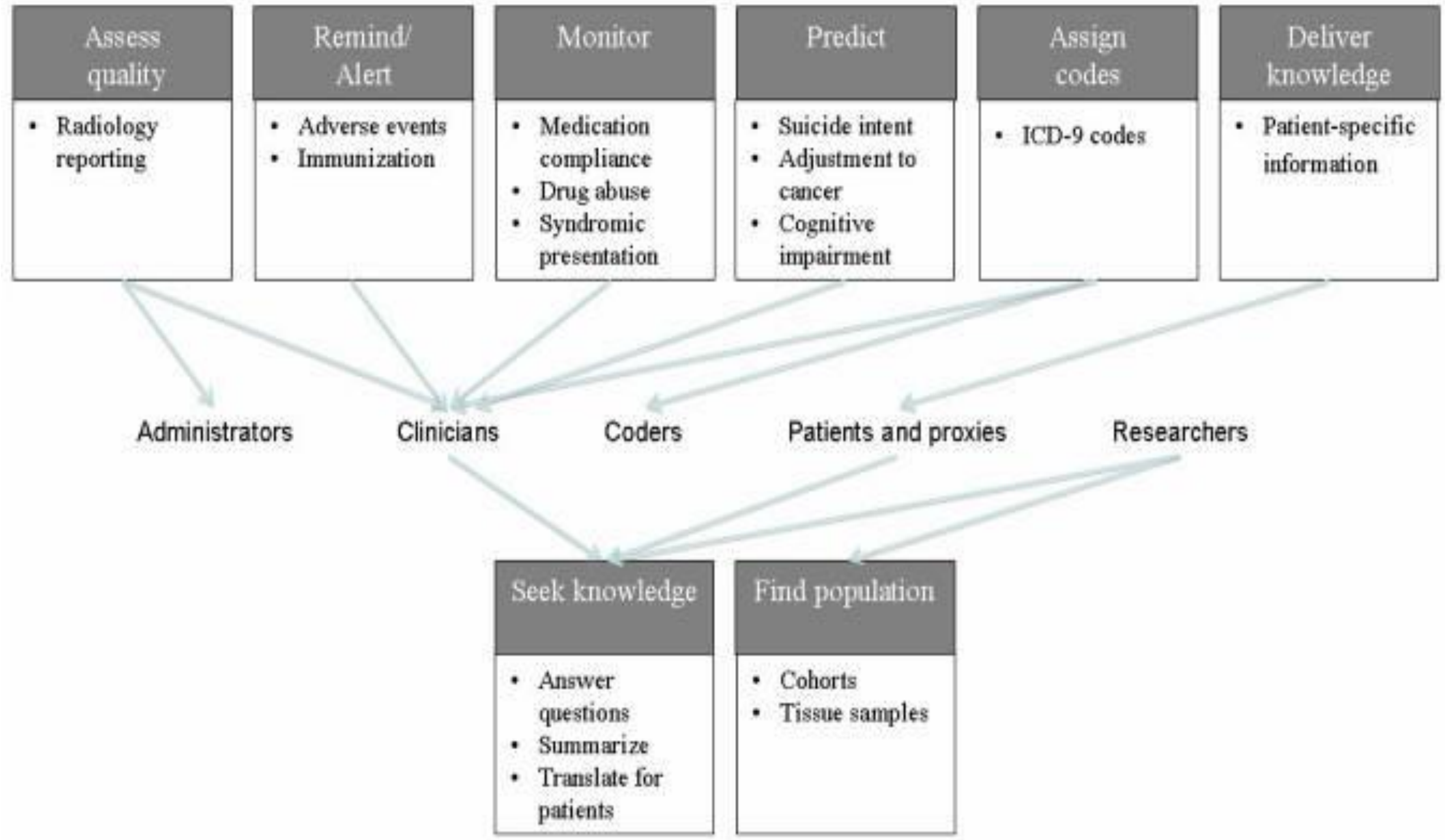


Figure: NLP in CDS

Source:
https://www.ncbi.nlm.nih.gov/core/lw/2.0/html/tileshop_pmc/tileshop_pmc_inline.html?title=Click%20on%20image%20to%20zoom&p=PMC3&id=2757540_nihms145183f1.jpg

NLP for CDS scope

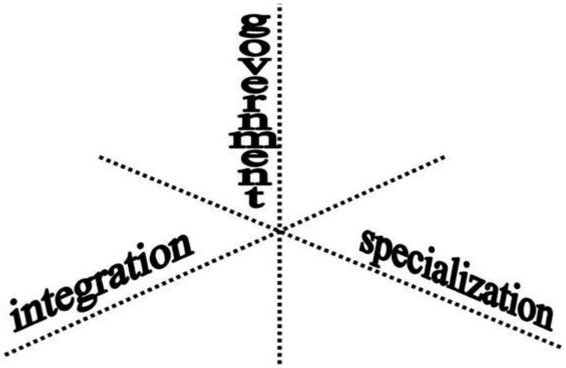


Figure: NLP-CDS Trifecta

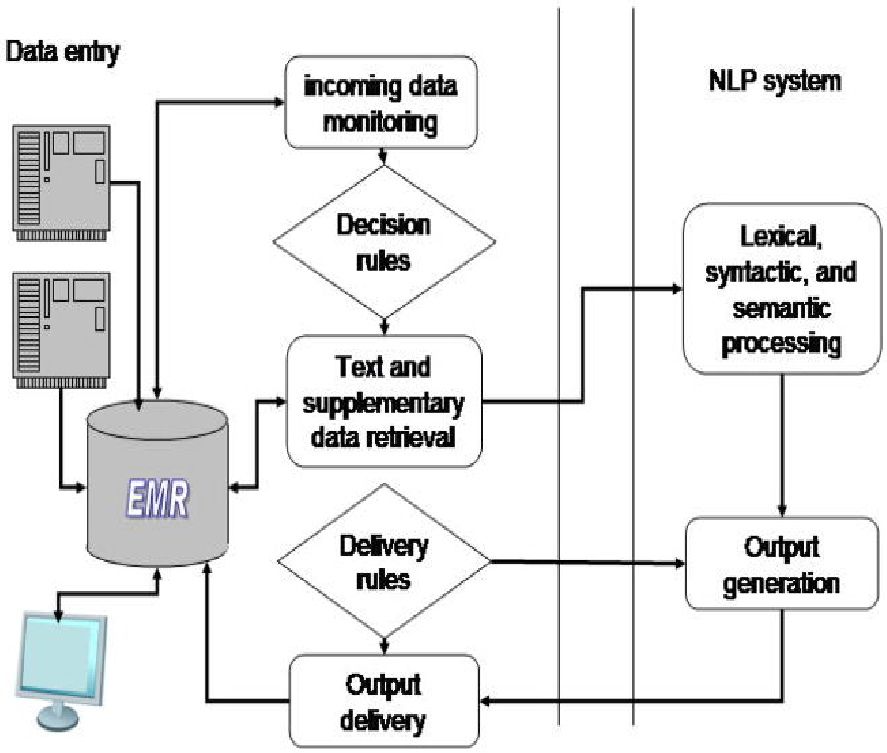


Figure: Specific NLP-CDS

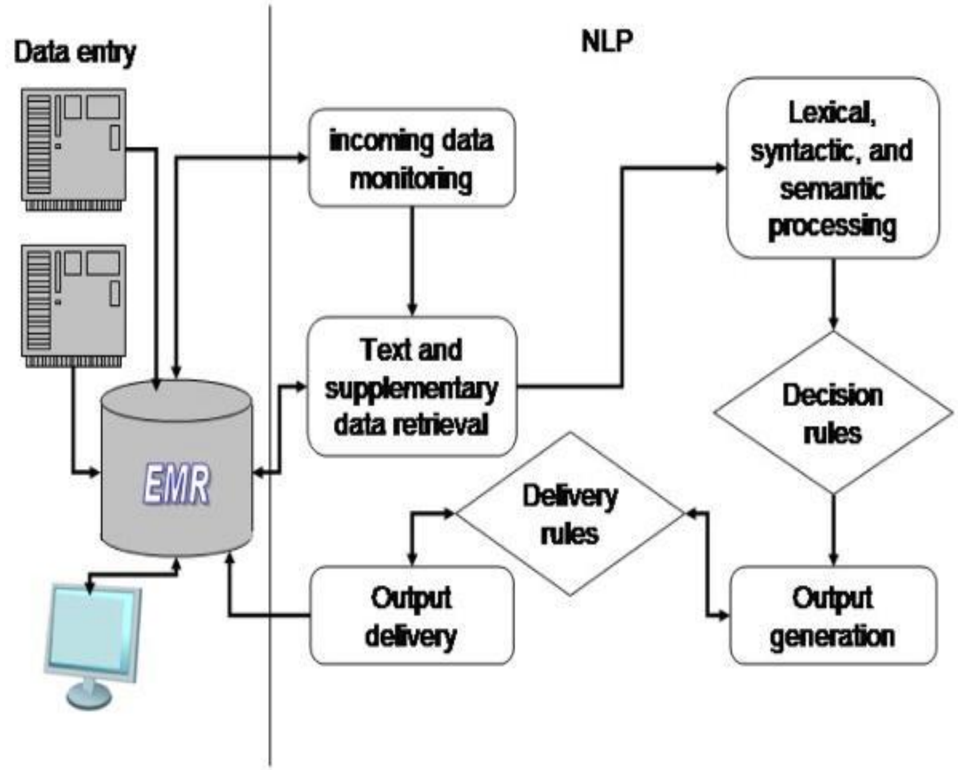


Figure: Generic NLP-CDS

Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2757540/

NLP models

- Generic model:
  - Modules ➔ Already customized, Predefined workflow structure.
- Specialized model:
  - Specific task ➔ Controls information flow
- Coupled Model:
  - Invoked by HER, Identify the presence of a disease, Create corresponding codes.

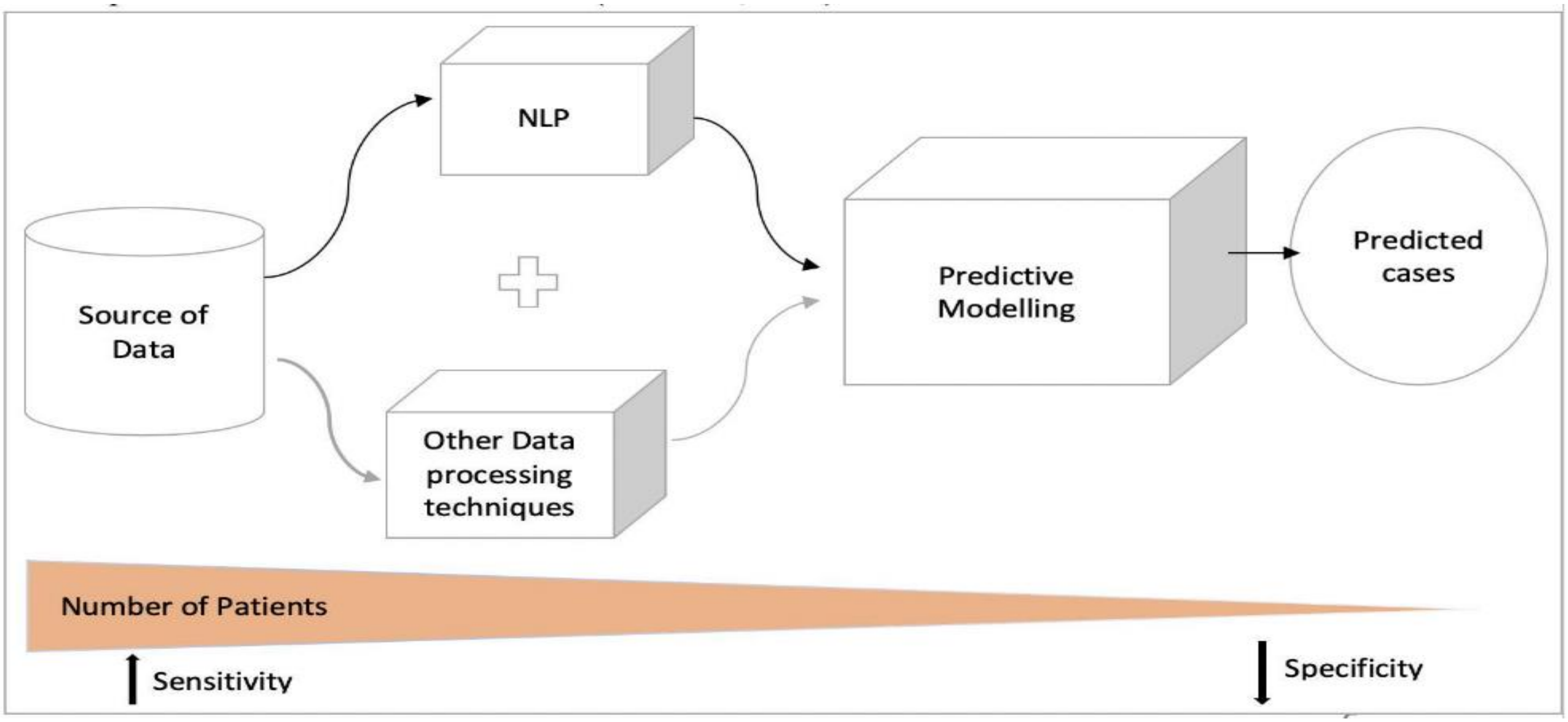


Figure: Generalization vs Specification

Source: https://ukdiss.com/examples/natural-language-processing.php

Building blocks of NLP - CDS

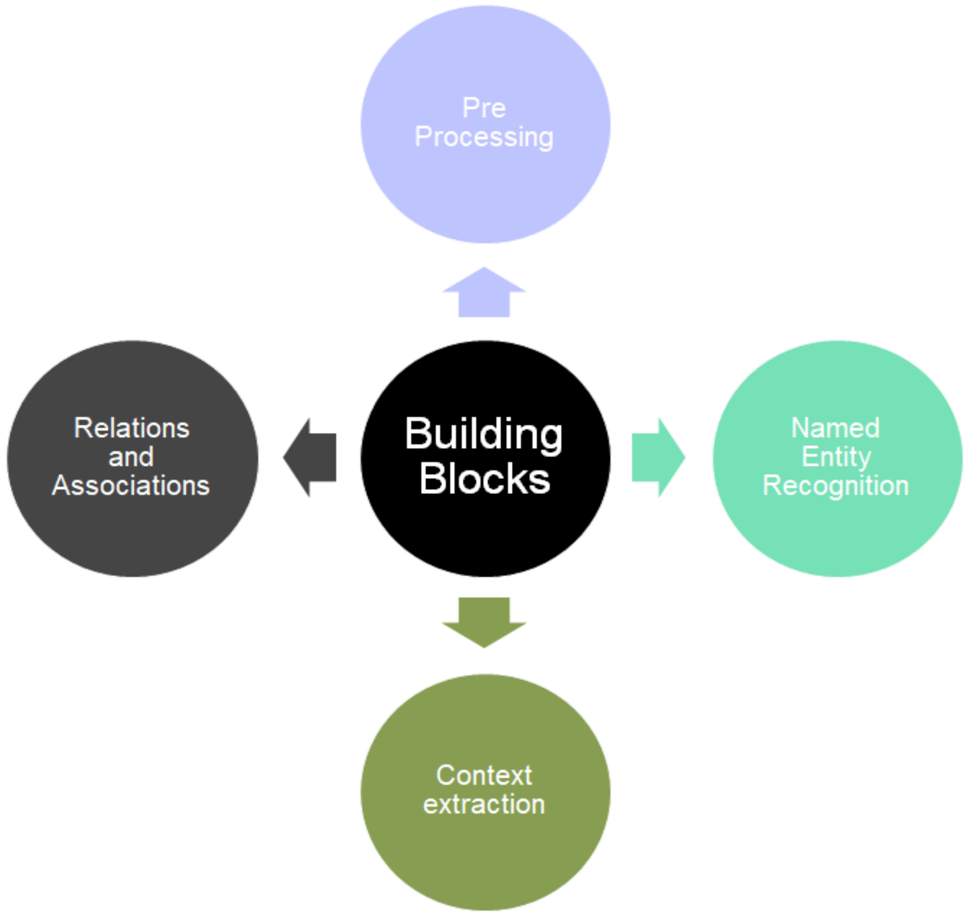


Figure: Building Blocks

Data based evidence collection: Summarization

Study	Design	Sample	Cognitive measures	Summary of results
Hajjar et al. (2009)[54] *USA*	Cross-sectional	*N*=580 community-dwelling elders (10% in low performance group had HF) Mean age 77.8 HF diagnosis: self-report	HVLT – Revised TMT-A TMT-B *Scores adjusted for age, education and race*	A group of elderly individuals with a phenotype of slow gait speed, greater depressive symptoms and worse executive function exists Elevated LDL is associated with being a member of this group (OR 1.01, 95% CI 1.00–1.02)
Kerola et al. (2010)[64] *Level of evidence: IV* *Finland*	Longitudinal descriptive cohort study	*N*=303 community-dwelling elders (Highest BNP group – 43% HF) Mean age 78.6–81.3 HF diagnosis: HF diagnosis using medical record	MMSE *Effect of age, sex, and education included in analysis*	Low HDL associated with lower MMSE at baseline (β=0.174, p=0.001)
Zuccalà et al. (1997)[32] *Level of evidence: VI* *Italy*	Cross-sectional descriptive	*N*=57 HF patients Mean age 77 HF diagnosis: systolic HF using echocardiogram and clinical criteria	MMSE *Effect of age and sex included in analysis*	Lower serum cholesterol associated with lower MMSE scores (r=0.30, p=0.02)

BNP: B natriuretic peptide; CI: confidence interval; HF: heart failure; LDL: low-density lipoprotein; MMSE: Mini Mental State Exam; OR: odds ratio; HVLT: Hopkins Verbal Learning Test; TMT: Trail Making Test

Figure: Sample summary

Source: https://www.researchgate.net/figure/Cognition-and-lipid-levels_tbl1_234099424

Applications of NLP in healthcare

Figure: Application Areas

Source: https://marutitech.com/nlp-in-healthcare/

Sentiment analysis and subjectivity

- Textual information:
  - Facts ➔ Objective expressions about entities.
  - Opinions ➔ Subjective expressions.

Figure: Sentiment Analysis

Source: https://medium.com/@tomyuz/a-sentiment-analysis-approach-to-predicting-stock-returns-d5ca8b75a42

Difficulties in sentiment analysis (1 of 2)

- Sentiment analysis ➔ Study of opinions, sentiments and emotions.
- Positive opinion & Negative opinions.
- Object o ➔ Entity ➔ o: (T, A).
- Opinion ➔ Feature f of an object O ➔ Positive or negative opinion on f.
- Feature f ➔ Explicit feature ➔ neither f nor any of its synonyms appear in s.
- Implicit feature ➔f is implied.

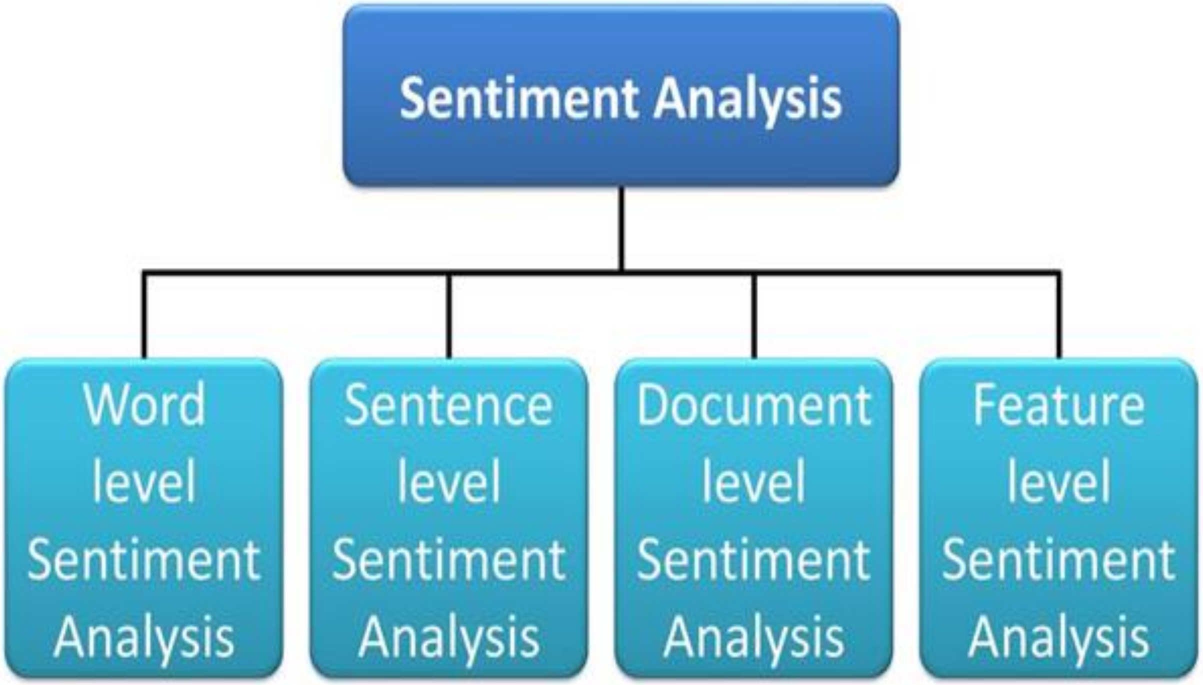


Figure: Levels of Analysis

Source: https://slideplayer.com/slide/12541406/75/images/32/Levels+of+Sentiment+Analysis.jpg

Difficulties in sentiment analysis (2 of 2)

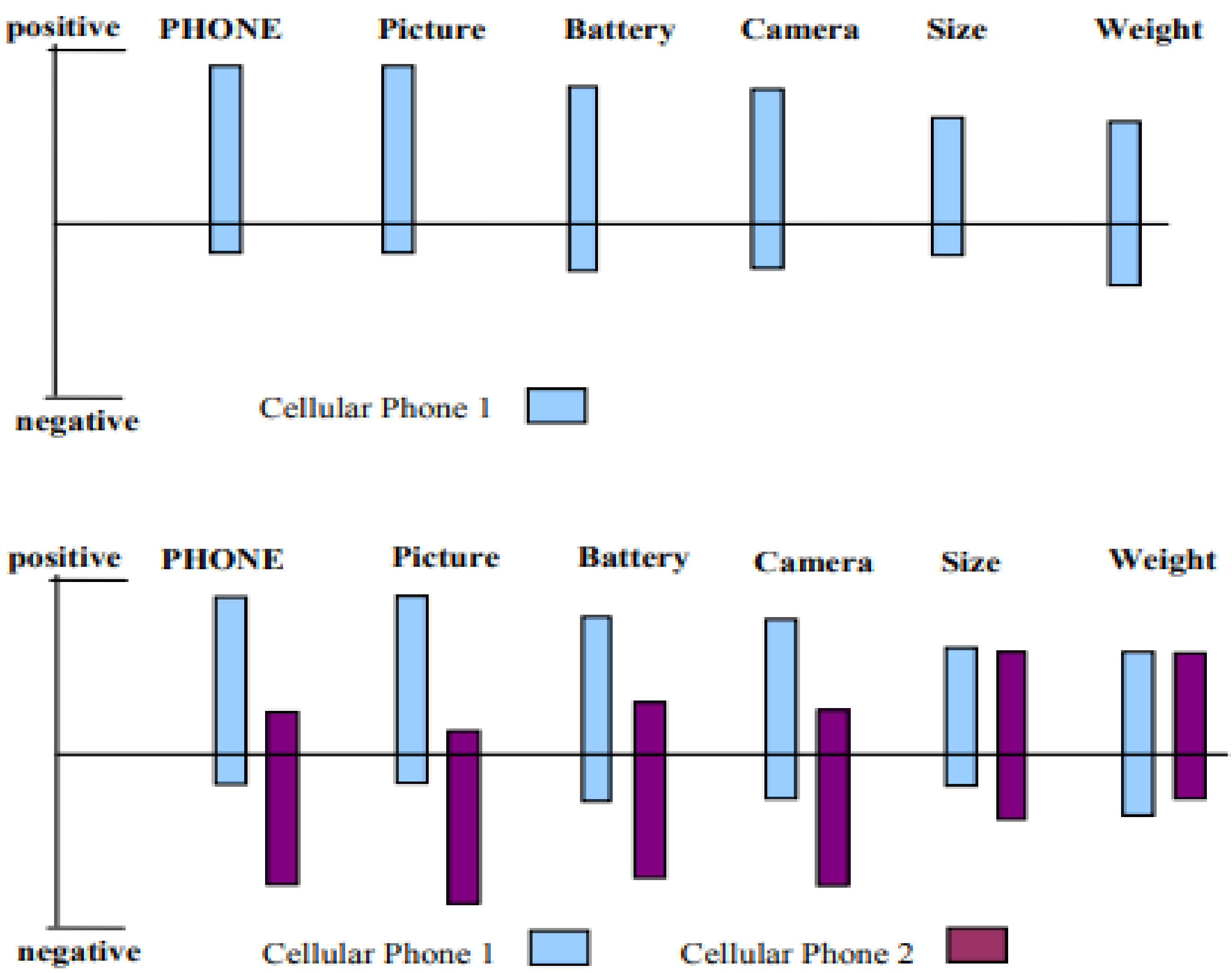


Figure: Opinion based Analysis

Source: https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf

Document level sentiment classification

	First word	Second word	Third word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Figure: Tagged Expression Sample

$$PMI(term_1, term_2) = \log_2\left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)}\right).$$

Figure: Input Term

Source: https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf

$$oo(phrase) = PMI(phrase, \text{“excellent”}) - PMI(phrase, \text{“poor”}).$$

Figure: Output Term

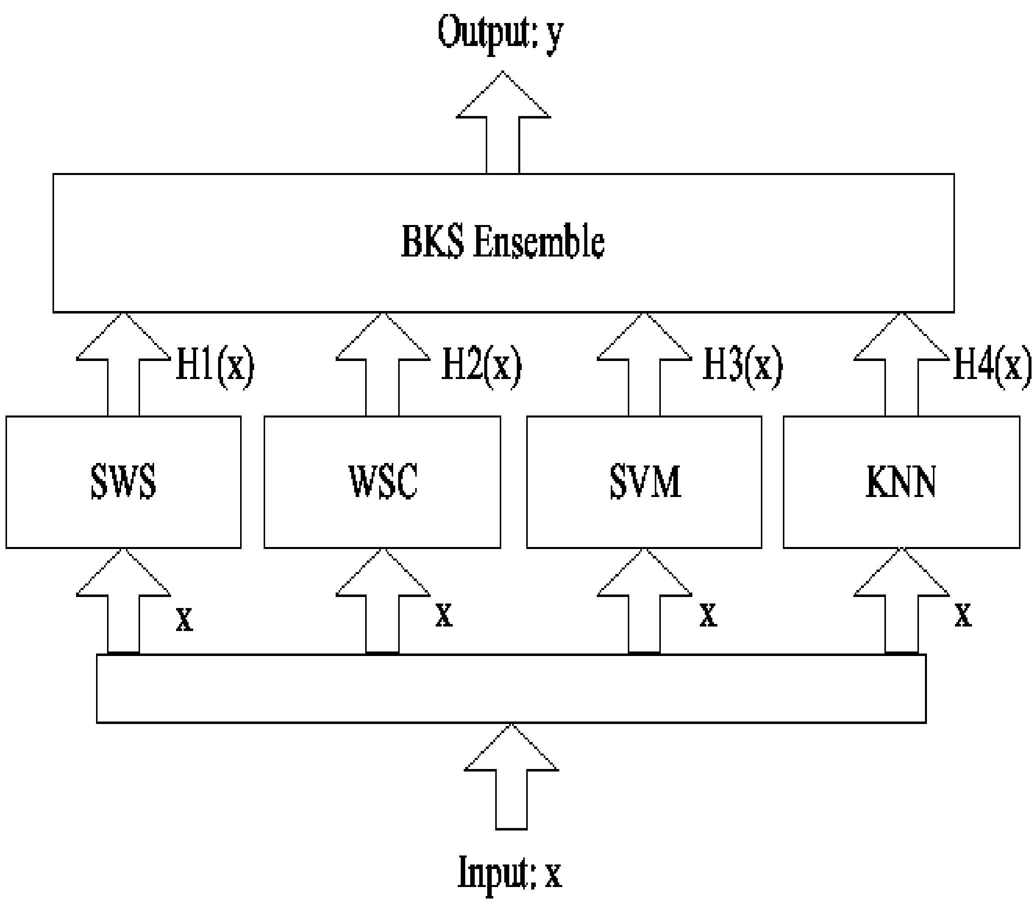


Figure: Doc level Representation

Source: https://www.semanticscholar.org/paper/Document-Level-Sentiment-Classification-Based-on-Zhang-Miao/3f9042aaefb517b7bf8b8f44b9db92d6d314a252/figure/3

Sentence level sentiment classification

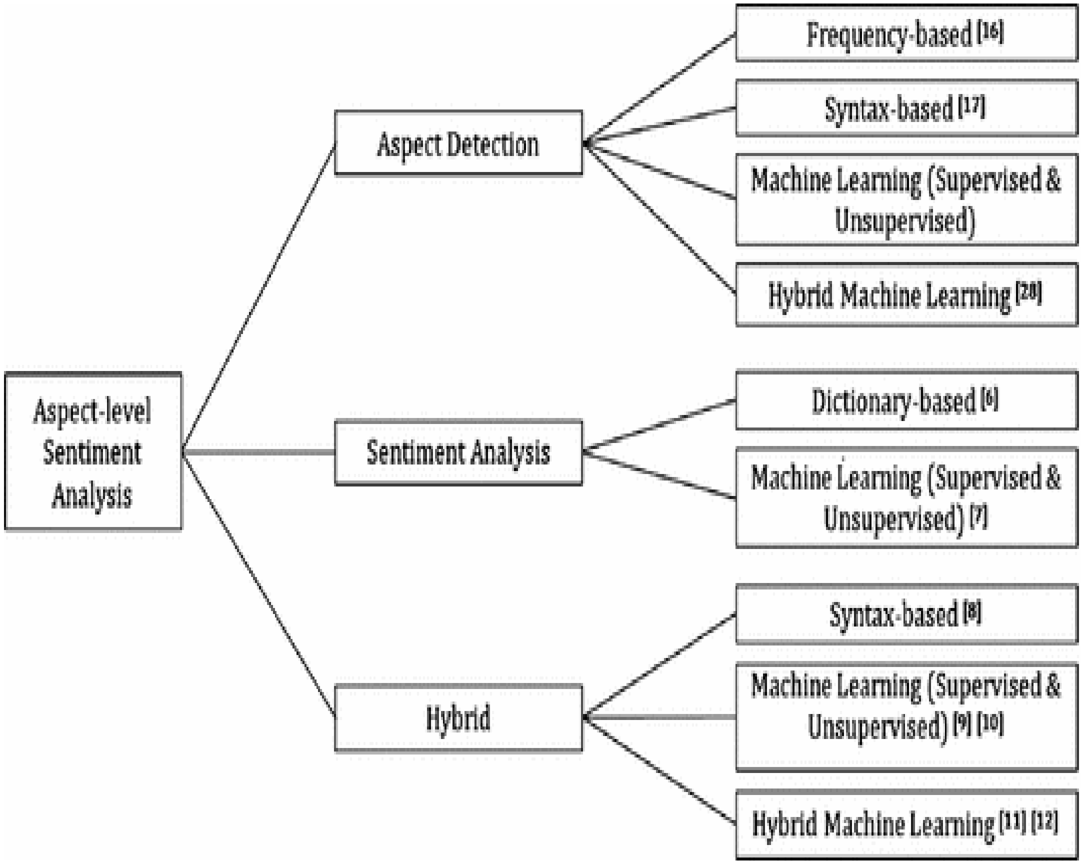


Figure: Sentence Level Sentiment Classification

Source: https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf

Syntactic template	Example pattern
<subj> passive-verb	<subj> was satisfied
<subj> active-verb	<subj> complained
active-verb <dobj>	endorsed <dobj>
noun aux <dobj>	fact is <dobj>
passive-verb prep <np>	was worried about <np>

Figure: Pattern

Source: https://link.springer.com/chapter/10.1007/978-981-13-1610-4_23

Lexicon (1 of 2)

- Base type:
  - Core words of opinion. Example: Good, beautiful, bad etc.
- Comparative type:
  - Comparison words of opinion Example: Better, worse, more important etc.

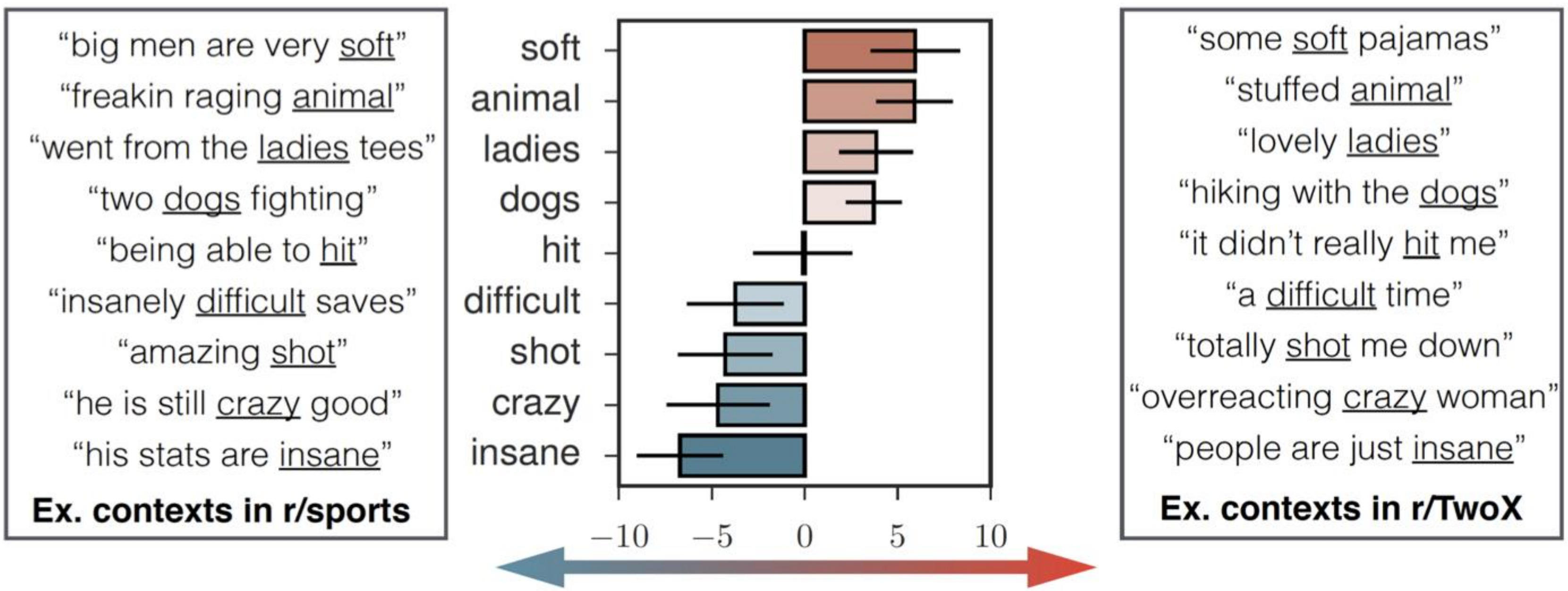


Figure: Domain specific Approach

Source: https://nlp.stanford.edu/projects/socialsent/

Lexicon (2 of 2)

Lexicon	Positive Words	Negative Words
Simplest (SM)	good	bad
Simple List (SL)	good, awesome, great, fantastic, wonderful	bad, terrible, worst, sucks, awful, dumb
Simple List Plus (SL+)	good, awesome, great, fantastic, wonderful, best, love, excellent	bad, terrible, worst, sucks, awful, dumb, waist, boring, worse
Past and Future (PF)	will, has, must, is	was, would, had, were
Past and Future Plus (PF+)	will, has, must, is, good, awesome, great, fantastic, wonderful, best, love, excellent	was, would, had, were, bad, terrible, worst, sucks, awful, dumb, waist, boring, worse
Bing Liu	2006 words	4783 words
AFINN-96	516 words	965 words
AFINN-111	878 words	1599 words
enchantedlearning.com	266 words	225 words
MPAA	2721 words	4915 words
NRC Emotion	2312 words	3324 words

Figure: Corpus collection of opinion words

Source: https://www.researchgate.net/figure/Examples-of-sentiment-lexicons_tbl1_288488744

Feature-based sentiment analysis (1 of 2)

Example Review:

Pros: Great photos, easy to use, very small

Cons: Battery usage; included memory is stingy.

I had never used a digital camera prior to purchasing this Canon A70. I have always used a SLR

Figure: Sample Review

great photos <photo>

easy to use <use>

very small <small>? <size>

battery usage <battery>

included memory is stingy <memory>

Simple Sequence: <{included, VB}{memory, NN}{is, VB}{stingy, JJ}>

Labelled Sequence: <{included, VB}{$feature, NN}{is, VB}{stingy, JJ}>

Rule: <{easy, JJ }{to}{*, VB}>? <{easy, JJ }{to}{$feature, VB}>confidence = 90%

Figure: Segregated Opinion Words

Feature-based sentiment analysis (2 of 2)

- Step 2: Opinion identification.

- Example: “The picture quality of this camera is not great, but the battery life is long.”

- Opinion words and phrases count: “The picture quality of this camera is not great [+1], but the battery life is long [0]”.

- Handling negations: “The picture quality of this camera is not great [-1], but the battery life is long [0]”.

- Usage of but clauses: “The picture quality of this camera is not great [-1], but the battery life is long [+1]”.

- Aggregating opinions:

$$score(f_i, s) = \sum_{op_j \in s} \frac{op_j.so}{d(op_j, f_i)},$$

Opinion summarization

Figure: Summarization

Source: https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf

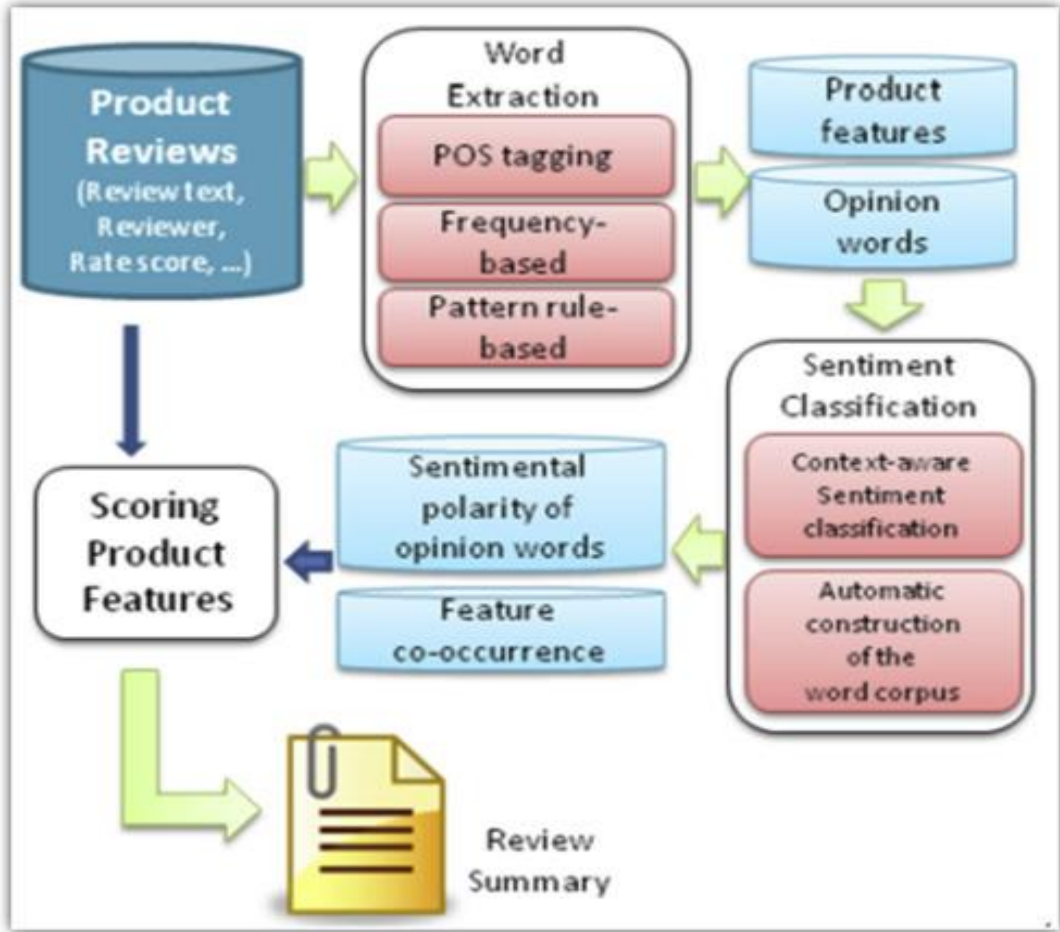


Figure: Sentiment Analysis as a Whole Process

Source: https://ars.els-cdn.com/content/image/1-s2.0-S2314728817300582-gr1.jpg

Self evaluation: Exercise 20

- To continue with the training, after learning the concepts of Information Retrieval and Question Answering in Natural Language Text Processing, it is time to write code to work with IR in NLP using the earlier topics implementing POS tagging, Tokenization and use it to compare similarities. It is instructed to utilize the concepts of reading data from files Tokenization, Word Similarity, POS tags and perform the following activity.

- You are instructed to write the following activities using Python code.

- Exercise 20: Python code to analyze the Sentiment based on subjects from a movie review Dataset.

Self evaluation: Exercise 21

- To continue with the training, after learning the concepts of Information Retrieval and Question Answering in Natural Language Text Processing, it is time to write code to work with IR in NLP using the earlier topics implementing POS tagging, Tokenization and use it to compare similarities. It is instructed to utilize the concepts of reading data from files Tokenization, Word Similarity, POS tags and perform the following activity.

- You are instructed to write the following activities using Python code.

- Exercise 21: Python code to analyze the Sentiment based on sentences from Twitter Samples.

Checkpoint (1 of 2)

Multiple choice questions:

1. Select correct statements related to the tasks of Sentiment analysis or opinion mining:
   a) Classifying the polarity of a given text at the document, sentence, or feature/aspect level
   b) Check, whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.
   c) Some Advanced tasks captures, "beyond polarity" sentiment classification looks
   d) All the above

2. NLP models are:
   a) Generic
   b) Specialized
   c) Coupled
   d) All the above

3. The auto tutor includes:
   a) Animated conversational agent
   b) Dialogue management
   c) Electronic documents
   d) All the above

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. Select correct statements related to the tasks of Sentiment analysis or opinion mining:
   a) Classifying the polarity of a given text at the document, sentence, or feature/aspect level
   b) Check, whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.
   c) Some Advanced tasks captures, "beyond polarity" sentiment classification looks
   d) All the above

2. NLP models are:
   a) Generic
   b) Specialized
   c) Coupled
   d) All the above

3. The auto tutor includes:
   a) Animated conversational agent
   b) Dialogue management
   c) Electronic documents
   d) All the above

Checkpoint (2 of 2)

Fill in the blanks:

1. Two main types of opinions are ____________and ____________________.
2. Opinion words are also called as _____________ words
3. Polysemy is defined as the coexistence of multiple meanings for a word or phrase in a text object. ________________________ model is the best choice to correct Polysemy.
4. While working with text data obtained from news sentences, which are structured in nature, ________________________ grammar-based text parsing techniques can be used for noun phrase detection, verb phrase detection, subject detection and object detection.

True or False:

1. Sentiment is the subset of emotion. True/False
2. Meaning of the text is important in text analysis for sentiments. True/False
3. Word2Vec model is a machine learning model used to create vector notations of text objects. Word2vec contains multiple deep neural networks. True/False

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. Two main types of opinions are **regular opinions** and **comparative opinions.**
2. Opinion words are also called as **polar** words.
3. Polysemy is defined as the coexistence of multiple meanings for a word or phrase in a text object. **Convolutional Neural Networks** model is the best choice to correct Polysemy.
4. While working with text data obtained from news sentences, which are structured in nature, **Dependency Parsing and Constituency Parsing** grammar-based text parsing techniques can be used for noun phrase detection, verb phrase detection, subject detection and object detection.

True or False:

1. Sentiment is the subset of emotion. **False**
2. Meaning of the text is important in text analysis for sentiments. **False**
3. Word2Vec model is a machine learning model used to create vector notations of text objects. Word2vec contains multiple deep neural networks. **False**

Question bank

Two mark questions:

1. How is the Layout of the document relevant in MM?
2. What are the features of why2atlas?
3. What is the core architecture of CDS based on?
4. What is feature based sentiment analysis?

Four mark questions:

1. How are CDS reports summarized?
2. Describe AUTOTUTOR architecture and process.
3. What is scripted dialogue and where to use it?
4. How to identify the meaning of text in MM?

Eight mark questions:

1. How is Layout and Meaning significant in Multimedia Presentation in NLP?
2. Write in detail the process steps involved in Sentiment Analysis with examples.

Unit summary

Having completed this unit, you should be able to:

- Gain knowledge on the process of Multimedia Presentation Generation
- Learn the concept of Language Interfaces for Intelligent Tutoring Systems
- Gain an insight into Argumentation for Healthcare Consumers
- Learn the concepts of Clinical Decision Support Systems
- Understand the core concepts of Sentiment Analysis and Subjectivity