



Ph.D. Defense: Semantic Annotation and Summarization of Biomedical Text

Lawrence H. Reeve

Committee Members:

- Dr. Hyoil Han**, IST (Advisor, Chair)
- Dr. Ari D. Brooks**, College of Medicine
- Dr. Xia Lin**, IST
- Dr. Il-Yeol Song**, IST
- Dr. Suk-Chung Yoon**, Widener University

Defense Overview

- Motivation / Research Questions
- Part I: Semantic Annotation
 - Related Work / Approach / Evaluation
- Part II: Text Summarization
 - Related Work / Approach / Evaluation
- Conclusions / Significance / Publications

Motivation - Annotation

- ‘Perfect Storm’ of data, resources and applications:
 - *Continuously updated data sources.*
 - National Library of Medicine’s PubMed:
 - > 16 million publications from 5,000 biomedicine-related journals
(United States National Library of Medicine, 2006a)
 - U.S. National Institutes of Health Clinical Trials database
 - > 13,500 clinical trials (United States National Library of Medicine, 2005a)
 - *New / Frequently-updated ontologies.*
 - UMLS Metathesaurus, NCI Thesaurus
 - *New ‘online’ applications using annotations.*
 - Our focus is on text summarization in general biomedical domain

Motivation - Summarization

- There exists no ‘ideal’ summary.
 - An author’s abstract is one view of an ideal summary
- The abstract may be missing content from the full-text.

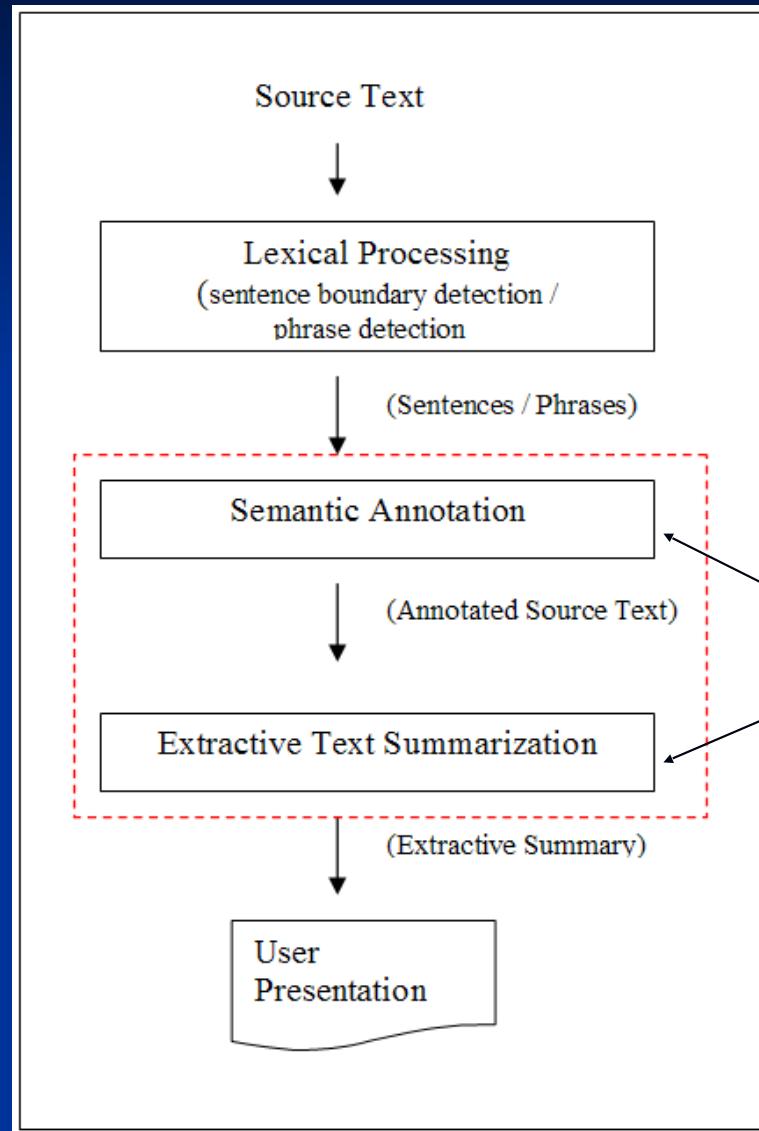
(Cohen & Hersh, 2005; National Institute of Health, 2005)
- The abstract may not be available.
- Customized summaries can be useful in question-answering systems where they provide personalized information.

(Mani & Bloedorn, 1998; Carbonell & Goldstein, 1998)

Research Questions

- Biomedical Semantic Annotation
 - Does the use of statistical language modeling methods improve biomedical semantic annotation performance over existing methods which use simple word metrics?
- Biomedical Text Summarization:
 - How well does a concept-based approach for biomedical text summarization perform as opposed to term-based summarization approaches?
 - Does adapting a lexical chaining approach for use with domain-specific concepts improve text summarization performance over existing approaches?
 - Does the use of frequency distribution of terms and/or concepts in a source text improve text summarization performance?

Integrated System



Focus of research

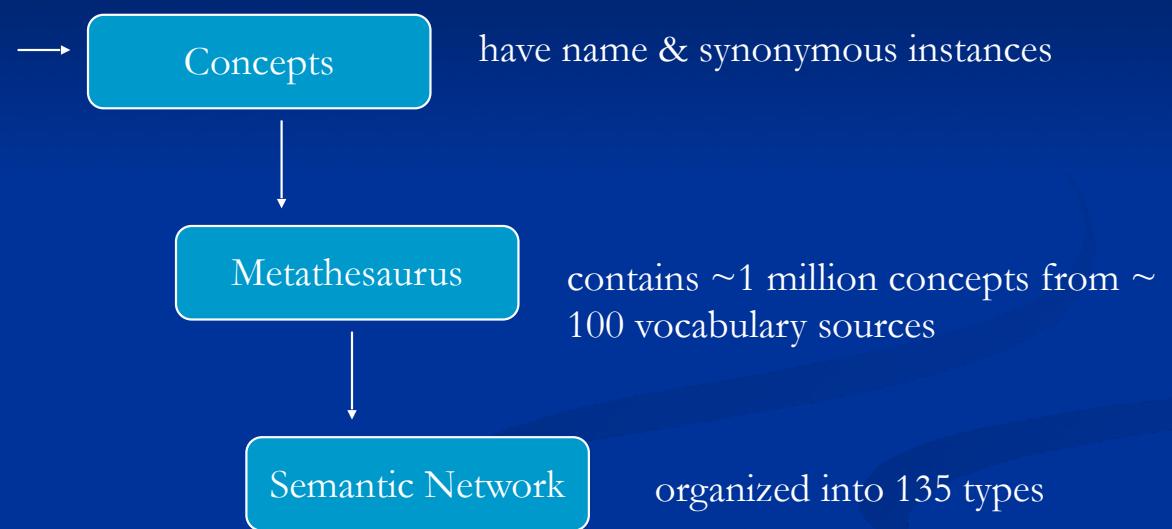
Biomedical Semantic Annotation

UMLS Metathesaurus

(Unified Medical Language System – U.S. National Library of Medicine)

Concept Name	Concept Instances
Multiple Myeloma	Multiple Myeloma
	Myeloma
	Plasma Cell Myeloma
	Myelomatosis
	Plasmacytic myeloma

(<http://umlsks.nlm.nih.gov>)



Semantic Annotation Example

- Given a source phrase: ‘Plasmacytic myeloma’
- Given concepts and concept instances:
- Map to UMLS concept: ‘Multiple Myeloma’

Concept Name	Concept Instances
Multiple Myeloma	Multiple Myeloma
	Myeloma
	Plasma Cell Myeloma
	Myelomatosis
	Plasmacytic myeloma

(<http://umlsks.nlm.nih.gov>)

Related Work – Biomedical Annotation Process

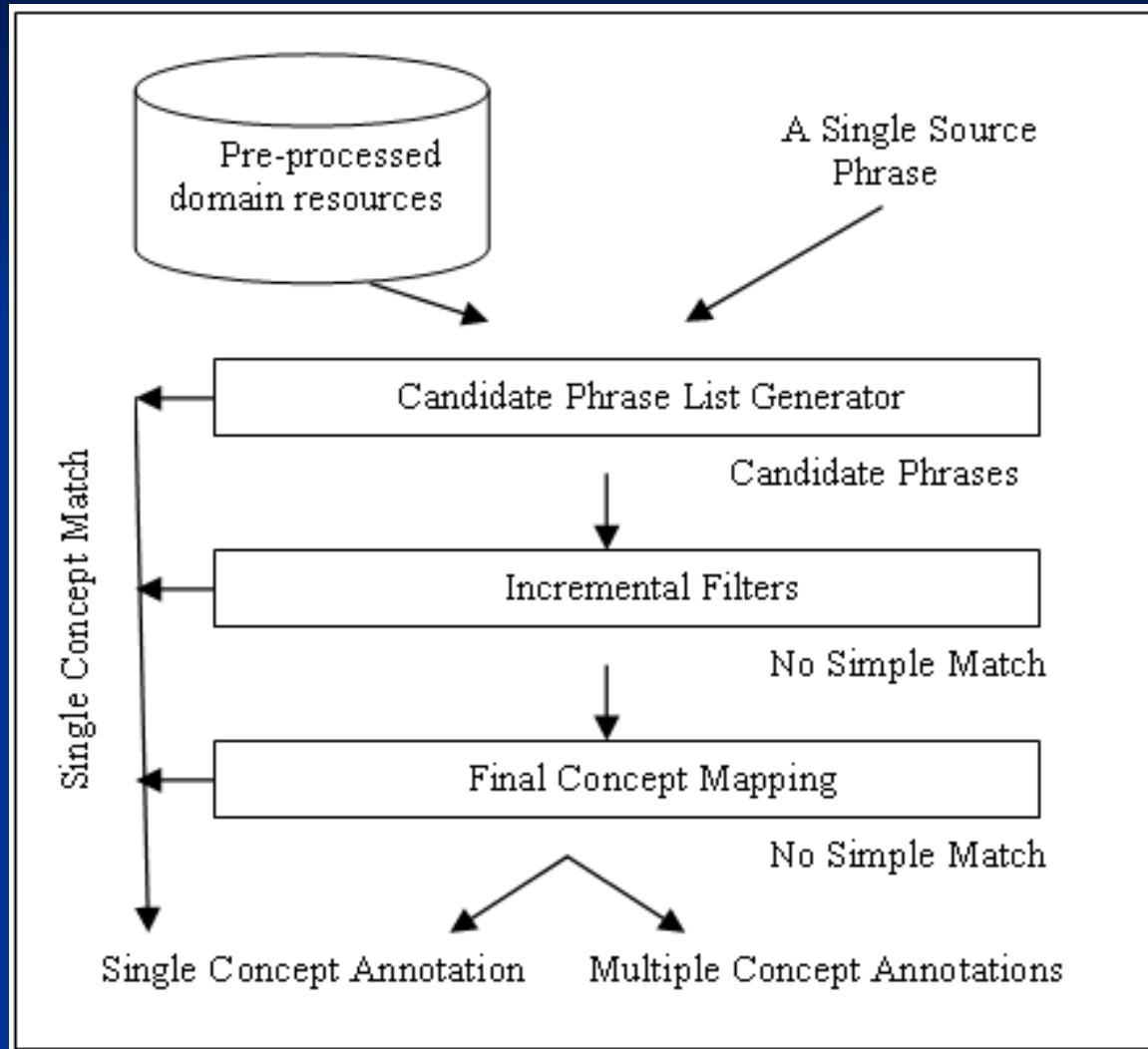
- Construct a unit of analysis (e.g., phrase, sentence) using subset of words in the source text.
- (Optional) Normalize the source text unit using UMLS (United States National Library of Medicine, 2005d) by:
 - (a) removing possessives,
 - (b) replacing punctuation with spaces,
 - (c) removing stop words,
 - (d) convert words to lower-case,
 - (e) breaking a string into constituent words, and
 - (f) sorting words into alphabetical order.
- For each word in the input phrase, build a set of all concepts containing each word.
- Find the intersection of the concept sets.
- (Optional) Find the best matching concept based on the common word membership between the source text and concept text.

Related Work – Biomedical Annotation Systems

System	Unit of Analysis	Phrase Identification Method	Mapping
SAPHIRE (W. R. Hersh, 1990)	Sentence	Text block	Simple, Partial
MetaMap Transfer (Aronson, 1996, 2001)	Phrase	NLP	Simple, Partial, Complex
SENSE (Ziemer & Bleich, 1997)	Phrase	User-specified Query	Simple
Concept Locator (P. Nadkarni et al., 2001)	Phrase	NLP Heuristics	Simple, Partial
Dynamic Taxonomy (Wollersheim et al., 2002)	Phrase	Moving Window	Simple
PhraseX (Srinivasan et al., 2002)	Phrase	NLP	Simple
KnowledgeMap (J. C. Denny et al., 2003)	Phrase	NLP	Simple, Partial
IndexFinder (Zou et al., 2003)	Unordered Terms	All words, excluding stop words	Simple, Partial

None: no match of terms; *Simple*: exact match; *Partial*: one or more terms do not match exactly; *Complex*: two or more sets of terms map to distinct concepts.

CONANN Architecture



CONANN = CONcept ANNNotator

Pre-processing: IPF Calculation

- Gives indication of word importance based on its usage within all UMLS concept phrases.
 - Pre-processing calculates IPF value of each word in UMLS phrases.

$$\text{Inverse Phrase Frequency} = \log \frac{N}{n_i}$$

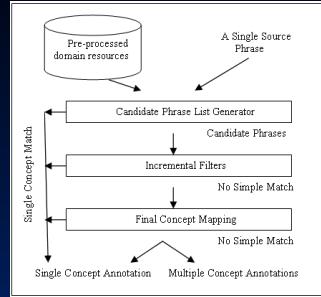
N is the number of UMLS phrases

i is a UMLS word.

n_i is the number of phrases word i appears in UMLS.

(Sparck Jones, 1972)

Coverage Filter



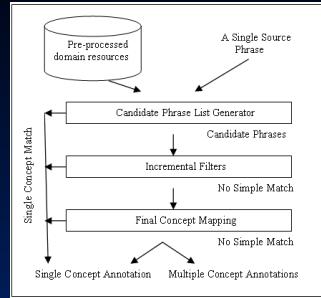
- How well do words overlap between the source phrase and the UMLS concept instance (phrase)?
 - Use IPF to find best overlap of most discriminative words.
 - Two approaches: *Naïve* and *Involvement*.

$$NaiveCandidatePhraseScore = \sum_{i=1}^N WordWeight_i$$

N : # words in common between a source phrase and a candidate phrase.

$WordWeight$ is either: 0 or 1 for binary weighting, or inverse phrase frequency value for IPF weighting

Coverage Filter



- Involvement Candidate Phrase Score:

$$Involvement = \frac{\sum_{i=1}^{NC} WordWeight_i}{|CandidatePhraseWords|} + \frac{\sum_{i=1}^{NS} WordWeight_i}{|SourcePhraseWords|}$$

NC: # words in candidate phrase that are in source phrase.

NS: # words in source phrase that are in candidate phrase.

WordWeight is either: 0 or 1 for binary weighting, or inverse phrase frequency value for IPF weighting

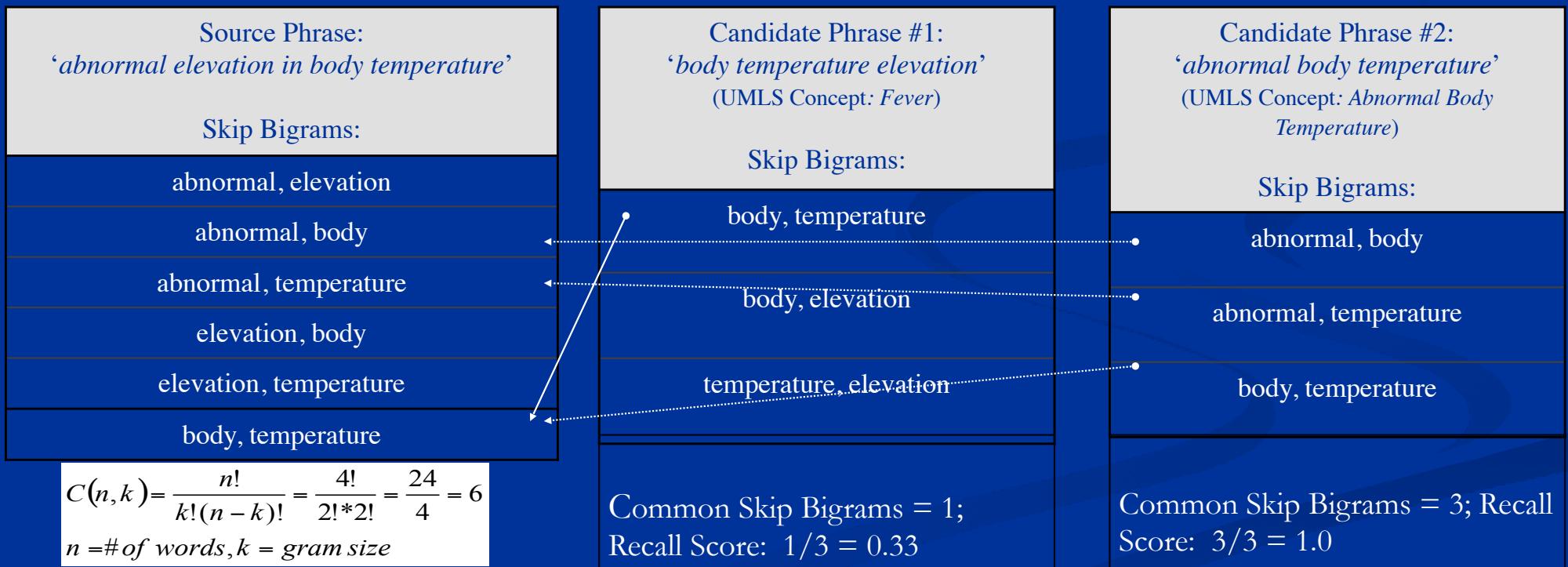
Aronson, A. R. (2001)

Coherence Filter

- Each candidate phrase is assigned a skip-bigram recall score:

$$Recall = \frac{CommonSkipBigrams(SourcePhrase, CandidatePhrase)}{CountSkipBigrams(CandidatePhrase)}$$

(C. Y. Lin & Och, 2004a)



Final Concept Mapping (Phrase Counting Method)

- 1) Candidate phrases first grouped by candidate concept.
- 2) Each candidate concept is then assigned a score based on how many candidate phrases belong to it.

Final Concept Mapping (Language Model method)

- 1) Generate a unigram language model for each concept using all concept instances within each concept. (pre-processing)
- 2) Get the probability that the candidate concept generated the source phrase:

$$P(SrcPhrase | M_{concept}) = \prod_{w \in SrcPhrase} ((1 - \lambda)P(w | M_{concept}) + \lambda P(w | M_{conceptCollection}))$$

- 3) Highest-probability candidate concept is then output as the best-matching concept for the source phrase.^(Manning, Raghavan, & Schütze, 2007)

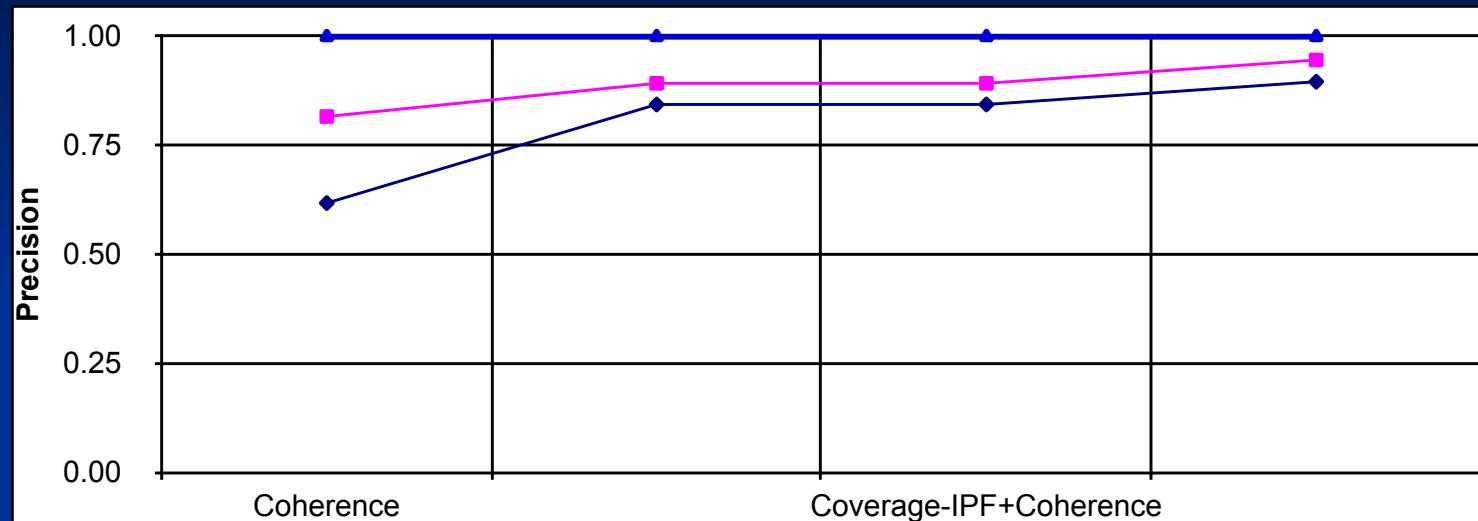
Semantic Annotation Evaluation

- Intrinsic
 - Evaluate speed and accuracy against a state-of-the-art biomedical concept annotator (MetaMap)
 - Execute MetaMap and CONANN annotators on identical corpus and compare output.
- Extrinsic
 - Show usefulness of concept output in text summarization task
 - Two summarizers use output of both MetaMap and CONANN to identify salient sentences.
 - Compare generated summaries using ROUGE tool (C. Lin, 2004)

Evaluation Corpus

- 24 random papers from a database of ~1,200 oncology randomized controlled trial publications Brooks, A. D., & Sulimanoff, I. (2002).
 - 1,600 distinct single-concept noun phrases from these papers
 - ROUGE tool suggests 24 samples with 3 model summaries (Lin, C. 2004)

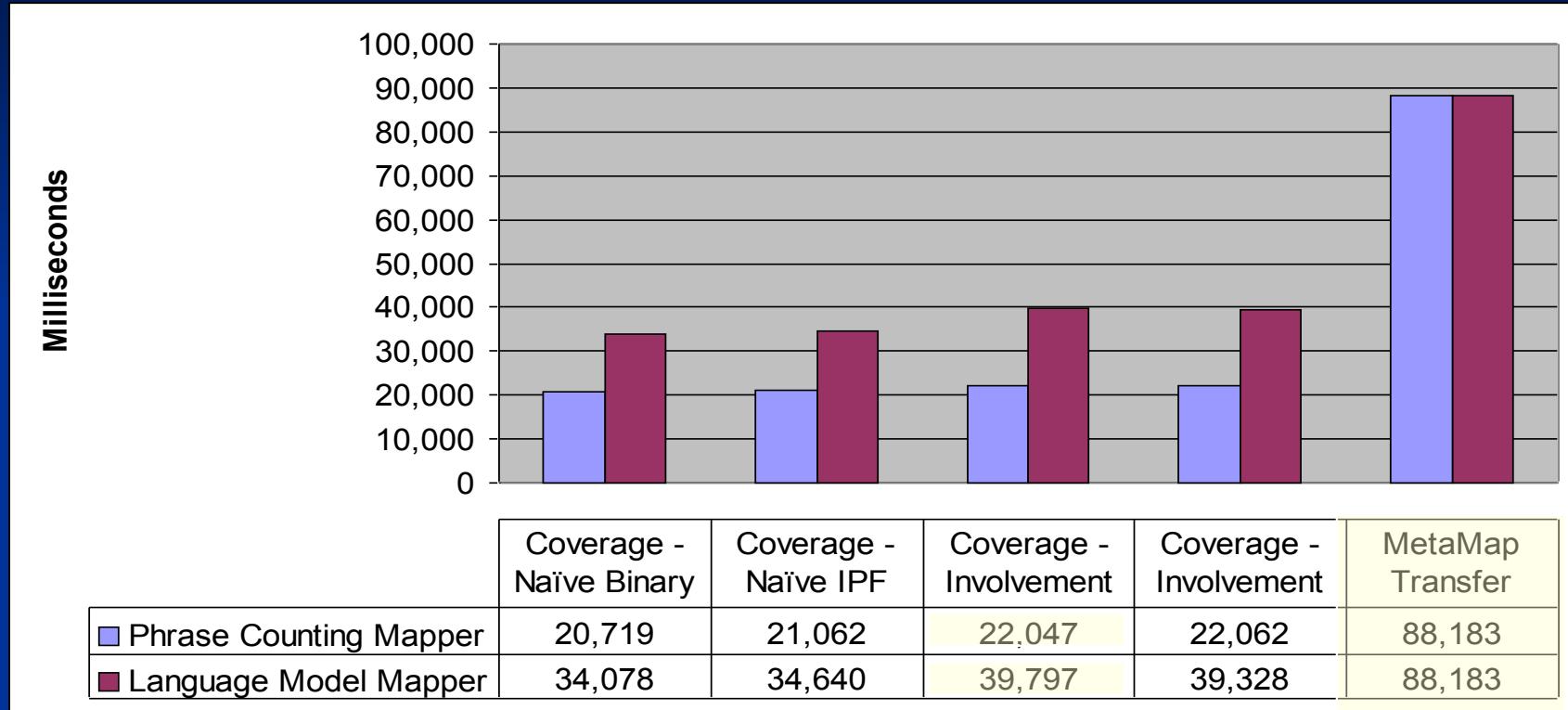
Annotation Results – Intrinsic



CONANN Precision Compared to MetaMap

- Coherence not effective alone; Coverage reasonable alone
- Order of filters matters
- Most best matching concepts are in top 5 CONANN output

Annotation Results – Intrinsic

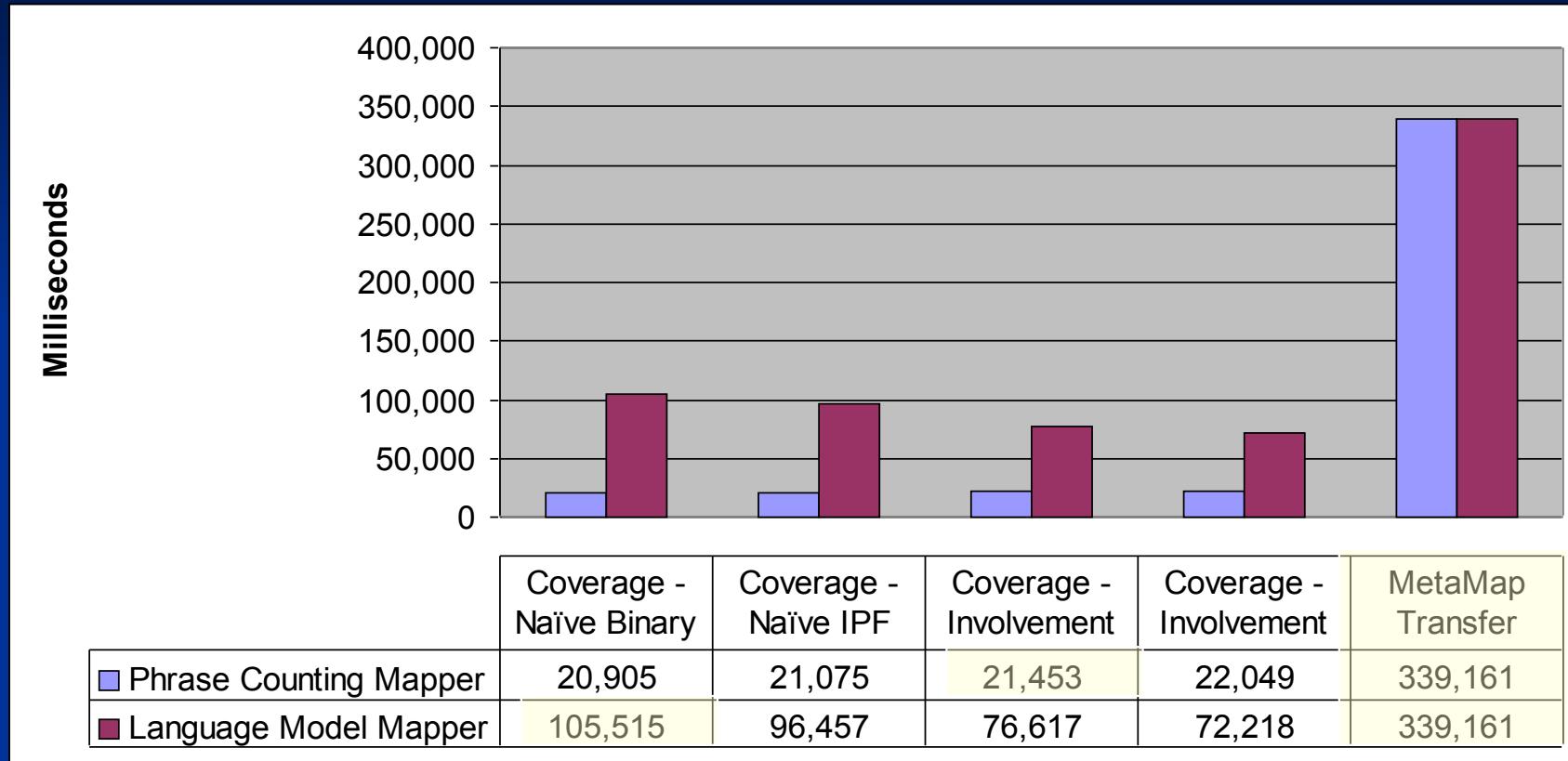


Annotator Initialization Time

-PCCM 4x faster, LMCM 2x faster: language models slow initialization time

-Keep external resources low

Annotation Results - Intrinsic



Total Annotation Time

PCCM 16x faster, LMCM 3x faster: language models require additional processing

Annotation Results – Intrinsic (Precision Using MetaMap as Baseline)

<i>Map</i>	<i>Filter</i>	<i>Exact Match</i>	<i>Top 5</i>
Phrase Counting Concept Mapper (PCCM)	Coverage - Involvement Binary	0.75	0.90
	Coverage - Involvement IPF	0.67	0.87
	Coverage - Naïve Binary	0.66	0.78
	Coverage - Naïve IPF	0.63	0.75
Language Model Concept Mapper (LMCM)	Coverage - Involvement Binary	0.85	0.93
	Coverage - Involvement IPF	0.81	0.91
	Coverage - Naïve Binary	0.76	0.86
	Coverage - Naïve IPF	0.74	0.83

- LMCM outperforms PCCM by about 0.10
- PCCM lower precision than LMCM, but faster
- Involvement outperforms naïve approach

Research Results – Biomedical Semantic Annotation

- Coverage filter works well.
 - Coherence not as useful.
- Phrase counting faster, but not as accurate.
 - Language model more accurate, but not as fast.
- Involvement outperforms Naïve scoring.
- Final mapping can be improved.
 - Best match is in top 5 concepts.

Biomedical Text Summarization:

Concept Chaining
Frequency Distribution
Hybrid

Related Work – Extractive Summarization

(Afantinos, Karkaletsis, & Stamatopoulos, 2005)

- Score each sentence of the text using some method which identifies important text.
- Extract the highest-scoring sentences.
- Order the sentences in their order of appearance.

Related Work – Biomedical Text Summarizers

Summarizer	Application	Source
PERSIVAL (PErsonalized Retrieval and Summarization of Images, Video and Language) (K. R. McKeown et al., 2001)	Personalized Summary	Physicians – clinical trials Laypeople - consumer health texts
Technical Article Summarizer (N. Elhadad, McKeown, Kaufman, & Jordan, 2005)	Personalized Summary	Patient record + clinical trials
Medical Text Indexer (National Library of Medicine, 2006)	Full-text Concept Indexing	PubMed articles
Medical Information Summarizer (Chen & Verma, 2006)	General Summary	PubMed
BioSquash (Shi et al., 2007)	Question-Answering	TREC Genomics
MiTAP (MITRE Text and Audio Processing) (Damianos et al., 2002)	Monitor biological threats	Multiple textual sources

Concept Chaining Summarizer

Concept Chaining Process

- 1) Chain together related concepts using UMLS Semantic Network.
- 2) Identify strongest chains & strongest concepts within the strongest chains.
- 3) Score each sentence based on number of strong concepts it contains.
- 4) Extract highest-scoring sentences.

Strong Chain Identification

- The strongest chain(s) indicate the set of concepts most representative of the source text.
- 1) Compute Chain score:
 - Uses reiteration and length
- 2) Find the strongest chains based chain scores:

$$\text{Score}(\text{Chain}) = \text{Frequency of most frequent concept} * \text{number of distinct concepts}$$

$$\text{Strong}(\text{Chain}) = \text{Score}(\text{Chain}) > (\text{Average}(\text{Scores}) + 2 * \text{StandardDeviation}(\text{Scores}))$$

(Barzilay & Elhadad, 1997)

Concept Frequency Distribution Summarizer

Concept Frequency Distribution Summarization Algorithm (FreqDist)

1. Build frequency model of source (full-text).
2. Iteratively select sentences from full-text:
 - Take a sentence, add it to current summary to form candidate summary.
 - Compare candidate summary distribution model for similarity to full-text distribution model.
 - Select sentence which, when added to summary, best models the full-text.
3. Repeat #2 until desired summary size reached.

Concept Distribution Similarity

- Problem: how to determine similarity between system-generated summary and full-text?
- Model system-generated summary and full-text as vector of concepts.
- Evaluate five similarity functions for the vectors:
 - *Cosine*: Calculate angle between the two vectors
 - *Dice*: Consider vector membership commonality
 - *Euclidean Distance*: Sum of squared distances
 - *Unit Item Frequency*: Fast simulation of cosine
 - *Vector Subtraction*: Subtract the two vectors

Example: FreqDistSumm at 5%

(Rose, P. G., et al., 1999)

- 1) Cisplatin is said to be the most active anticancer drug in cervical cancer, although 100 mg/m did not significantly improve the complete response (CR) rate compared prospectively with 50 mg/m (12.7 v 10%, respectively) and there was no appreciable difference in response duration, progression-free interval (PFI), or survival.
- 2) Thus, there is no convincing reason to use a cisplatin dose higher than 50 mg/m .
- 3) However, shortly after activation, the study was amended so that the serum creatinine concentration was to be within the institution's normal limits and the serum albumin level was to be ≥ 3 g/dL .
- 4) Ineligible patients included those with cervical neoplasms other than squamous cell carcinoma or with nonmeasurable cervical cancer, WBC count less than 4,000/uL and/or platelet count less than 100,000/uL, abnormal liver function (bilirubin, AST, or alkaline phosphatase level $>$ two times normal not related to the cancer), bilateral hydronephrosis, GOG PS 3 or 4, past or concomitant malignancy other than skin (excluding melanoma), prior therapy with cytotoxic drugs except when used as a radiation sensitizer, radiationtherapy within 3 weeks of entry, lesions measurable only by ultrasound, or pregnancy or lactation .

Example: FreqDistSumm at 5%

(Rose, P. G., et al., 1999)

- 5) Patients were prospectively stratified according to whether they had received prior radiation-sensitizer treatment (hydroxyurea, cisplatin, or fluorouracil) and by PS, and were then centrally randomized with equal probability to receive (1) cisplatin 50 mg/m with appropriate hydration every 3 weeks for a maximum of six courses, or (2) cisplatin 50 mg/m on day 1 plus mitolactol (180 mg/m orally for 5 days) on days 2 through 6 every 3 weeks, or (3) cisplatin 50mg/m plus ifosfamide 5.0 g/m over 24 hours plus mesna 6 g/m given concurrently with ifosfamide and for 12 hours after, every 3 weeks, again for a maximum of six courses .
- 6) Between June 1990 and January 1994, 454 women entered the study of whom 16 were ineligible (wrong stage, n = 2, wrong cell type, n = 9, wrong primary tumor, n = 2, prior chemotherapy, n = 2, second primary tumor, n = 1), which left 438 eligible patients (including 10 who received no drug, all were included in the intent to-treat analysis) .
- 7) The results of this trial in cervix cancer reflect a common set of findings in combination chemotherapy of advanced solid tumors, namely, a higher response rate (but not a high CR rate) with a combination compared with single-agent therapy at the cost of more toxicity and no survival benefit.

Text Summarization Evaluation

- Use ROUGE (Recall Oriented Understudy for Gisting Evaluation) (Lin & Hovy, 2003)
 - Recall score based on n-gram overlap between system-generated summary and model summaries:
- ROUGE-2
 - Bigram co-occurrence between system-summary and model summaries^(Lin, C, 2004)
- ROUGE-SU4
 - Skip bigram, with distance of 4, co-occurrence
 - (bigrams with no more than 4 intervening words)
- Same metrics are used in Document Understanding Conferences

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{gram_n \in S} Count(gram_n)}$$

Evaluation - Summarizers

- Baseline
 - LEAD: first 20% of sentences in text
 - RANDOM: randomly select 20% of sentences in text
- Frequency-based:
 - AutoSummarize (part of Microsoft Word) (Microsoft Corporation, 2002)
 - Open Text Summarizer (OTS) (Rotem, 2003)
- Multiple Feature:
 - MEAD: (uses sentence position, sentence length, clustering) (Radev et al., 2004)
 - SWESUM: (uses sentence position, higher weights for number values) (Dalianis, 2000)
- Reduce Information Redundancy
 - Lemur MMR (The Lemur Project, 2006)
 - SumBasic (frequency-based; probabilistic; state of the art) (A. Nenkova & Vanderwende, 2005)

Results - BioChainSumm

- All versions of BioChain outperformed all other summarizers except for Lemur and SumBasic.
 - Lemur and SumBasic perform redundancy removal
- Terms outperformed concepts.

BioChainSumm ROUGE-2

Summarizer	ROUGE-2
SumBasic-Term	0.11673
SumBasic-Concept	0.10940
Lemur-MMR	0.10708
BioChain-MostFrequentStrongChainConcept	0.10419
BioChain-MostFrequentStrongChainConcept-Position	0.10175
BioChain-AllStrongChainConcepts-Filtered	0.10043
BioChain-AllStrongChainConcepts-Position-Filtered	0.10043
BioChain-MostFrequentStrongChainConcept-Filtered	0.09868
BioChain-AllStrongChainConcepts	0.09708
BioChain-AllStrongChainConcepts-Position	0.09708
BioChain-MostFrequentStrongChainConcept-Position-Filtered	0.09660
MEAD	0.09254
Baseline-Random	0.08001
AutoSummarize	0.07977
SweSum	0.07513
OTS	0.07474
Baseline-Lead	0.07076

Results - FreqDistSumm

- FreqDistSumm outperformed all other summarizers.
- Dice similarity comparison was best.
- Terms slightly outperformed concepts.

FreqDistSumm ROUGE-2

Summarizer	ROUGE-2
FreqDist-Term_Dice	0.22176
FreqDist-Concept_Dice	0.21997
SumBasic-Term	0.21112
FreqDist-Term_UnitFrequency	0.20707
SumBasic-Concept	0.20034
FreqDist-Concept_Cosine	0.19932
FreqDist-Concept_UnitFrequency	0.19932
MEAD	0.17629
FreqDist-Term_Cosine	0.17358
Baseline-Random	0.16396
AutoSummarize	0.15171
SweSum	0.15115
Baseline-Lead	0.13953
FreqDist-Concept_VectorSubtraction	0.11435
FreqDist-Concept_Euclidean	0.09236
FreqDist-Term_Euclidean	0.07516
FreqDist-Term_VectorSubtraction	0.05742

Results - ChainFreqSumm

- ChainFreqSumm outperformed all other summarizers.
- Use of all strong chain was better than most frequent strong chain concept.
- Concepts outperformed terms (ROUGE-SU4).

ChainFreqSumm ROUGE-SU4

Summarizer	ROUGE-SU4 Score
ChainFreqSumm-AllStrongChainConcepts-Dice	0.22303
FreqDistSumm-Term-Dice	0.22176
FreqDistSumm-Concept-Dice	0.21997
SumBasic-Term	0.21112
ChainFreqSumm-MostFrequentStrongChainConcept-Dice	0.20158
SumBasic-Concept	0.20034
Lemur-MMR	0.19874
BioChainSumm-MostFrequentStrongChainConcept	0.19173
BioChainSumm-AllStrongChainConcepts	0.18557
Mead	0.17629
Baseline-Random	0.16396
MSWord	0.15171
SweSum	0.15115
OTS	0.14919
Baseline-Lead	0.13953

FreqDist ROUGE Score Significance

ROUGE-2

External Summarizers	ROUGE-2 Scores	FreqDist Summarizers
SumBasic-Term	0.12711	0.13323 Term
SumBasic-ConceptDUIST	0.11827	0.12080 ConceptMMTX
SumBasic-ConceptMMTX	0.10920	0.11943 ConceptDUIST
Lemur-MMR	0.09653	
Mead	0.09160	
Baseline-Random	0.07913	
MSWord	0.07849	
SweSum	0.07430	
OTS	0.07364	
Baseline-Lead	0.07112	

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
External Summarizers	10	0.91939	0.091939	0.000408
FreqDist Summarizers	3	0.37346	0.124486667	5.78E-05

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.002444655	1	0.002444655	7.096378	0.022033	4.844338
Within Groups	0.003789427	11	0.000344493			
Total	0.006234082	12				

$$F(10, 3) = 7.10, p < .05$$

ROUGE-SU4

External Summarizers	ROUGE-SU4 Scores	FreqDist Summarizers
SumBasic-Term	0.22169	0.23005 Term
SumBasic-ConceptDUIST	0.20894	0.21864 ConceptMMTX
SumBasic-ConceptMMTX	0.19868	0.21487 ConceptDUIST
Lemur-MMR	0.19737	
Mead	0.17368	
Baseline-Random	0.16100	
MSWord	0.14886	
SweSum	0.14848	
OTS	0.14646	
Baseline-Lead	0.13706	

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
External Summarizers	10	1.74222	0.174222	0.000914
FreqDist Summarizers	3	0.66356	0.221186667	6.25E-05

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.00509003	1	0.005090031	6.705222	0.025163	4.844338
Within Groups	0.00835026	11	0.000759114			
Total	0.01344029	12				

$$F(10, 3) = 6.71, p < .05$$

- FreqDist summarizers statistically significantly different in ROUGE scores than other summarizers.

ROUGE Scores at Various Summary Sizes

Summarizer	Score Type	1%	5%	10%	15%	20%
FreqDist-Concept	ROUGE-2	0.01561	0.05817	0.09063	0.10540	0.12069
FreqDist-Term	ROUGE-2	0.01561	0.05526	0.10289	0.11944	0.13241
SumBasic-Concept	ROUGE-2	0.02432	0.06128	0.08102	0.10303	0.11003
SumBasic-Term	ROUGE-2	0.06098	0.08532	0.10909	0.11800	0.12611
FreqDist-Concept	ROUGE-SU4	0.06766	0.13398	0.18256	0.20260	0.21851
FreqDist-Term-Dice	ROUGE-SU4	0.06766	0.12737	0.19341	0.21382	0.22941
SumBasic-Concept	ROUGE-SU4	0.07277	0.13387	0.15964	0.19106	0.19941
SumBasic-Term	ROUGE-SU4	0.12280	0.16383	0.19802	0.21173	0.22089

Starting at 10% level, summary size results in less dramatic score changes

Research Results – Biomedical Text Summarization

- Domain-specific summarizers outperform general-purpose summarizers.
- Concept Chaining is good at identifying important areas of text.
- Frequency distribution method controls redundancy.
- Concepts can used in place of terms for identifying important areas of text.
- Ideal summary size for clinical trial texts is around 10% of source text.

Annotation Results – Extrinsic (Phrase Counting Final Mapper)

<i>Summarizer</i>	<i>ROUGE-2 Score</i>	<i>ROUGE-SU4 Score</i>
FreqDist using MetaMap	0.12080	0.21864
FreqDist using CONANN (Coverage - Involvement Binary)	0.13018	0.22361
FreqDist using CONANN (Coverage - Naïve IPF*)	0.12897	0.22252
FreqDist using CONANN (Coverage - Naïve Binary)	0.12872	0.22199
FreqDist using CONANN (Coverage – Involvement IPF*)	0.12872	0.22199
SumBasic using MetaMap	0.11412	0.19868
SumBasic using CONANN (Coverage - Involvement Binary)	0.11834	0.21039
SumBasic using CONANN (Coverage – Involvement IPF*)	0.11827	0.20770
SumBasic using CONANN (Coverage - Naïve IPF*)	0.11702	0.20670
SumBasic using CONANN (Coverage – Naïve Binary*)	0.11210	0.20191

CONNAN with IPF word weights in word coverage filtering outperforms MetaMap annotations from 2.5% to 6.0%

Annotation Results – Extrinsic (Language Model Final Mapper)

Summarizer	ROUGE-2 Score	ROUGE-SU4 Score
FreqDist using MetaMap	0.12080	0.21864
FreqDist using CONNAN (Coverage - Naïve Binary)	0.12897	0.22252
FreqDist using CONNAN (Coverage - Naïve IPF*)	0.12897	0.22252
FreqDist using CONNAN (Coverage – Involvement Binary)	0.12897	0.22292
FreqDist using CONNAN (Coverage – Involvement IPF*)	0.12897	0.22292
SumBasic using MetaMap	0.10920	0.19868
SumBasic using CONANN (Coverage - Naïve Binary)	0.12028	0.21212
SumBasic using CONANN (Coverage - Naïve IPF*)	0.11614	0.20794
SumBasic using CONANN (Coverage - Involvement Binary)	0.11839	0.21053
SumBasic using CONANN (Coverage - Involvement IPF*)	0.11839	0.21053

CONNAN with IPF word weights in word coverage filtering outperforms MetaMap annotations from 1.7% to 7.8%

Conclusions

Conclusions

- Biomedical Semantic Annotation:
 - Annotation time can be significantly decreased (by 3x-16x) with precision competitive with a state-of-the-art annotator.
 - Support texts unknown to the system ahead-of-time
 - Supports constantly changing ontologies
 - Application of statistical language modeling improves performance over heuristic methods and simple word metrics.
 - Considers all concept instances, rather than individual concept instances.
 - Provide multiple performance variations.
 - Phrase counting: 16x faster with lower precision
 - Concept language models: 3x faster with higher precision
 - All variations can be used for online applications (seconds vs. minutes)

Conclusions

- Biomedical Text Summarization
 - Domain-specific concepts can be used in place of terms to identify salient areas of text.
 - This is a new approach in biomedical text summarization.
 - Developed two concept-based algorithms:
 - Concept Chaining finds salient domain-specific information.
 - Others linked related terms.
 - Frequency Distribution controls redundancy.
 - This is a new approach in text summarization.
 - Combining concept chaining and frequency distributions provides best results.
 - Domain-specific summarizers outperform general-purpose summarizers.
 - The ideal biomedical text summary size is about 10%.
 - This is the first study in biomedical text summarization.

Significance

- **Biomedical Text Summarization:**
 - Pioneered the use of domain-specific concepts in place of terms for biomedical text summarization.
 - Concepts overcome biomedical language variation.
 - Applications: Summarization Personalization, Multiple document summarization, Q&A systems
- **Biomedical Semantic Annotation:**
 - Applied statistical language modeling from speech recognition field to biomedical semantic annotation in natural language processing field.
 - Integrated knowledge from one disciplinary area to another.
 - Annotation time can be significantly decreased (by 3x-16x) with precision competitive with a state-of-the-art annotator.
 - Support texts unknown to the system ahead-of-time as well as constantly changing ontologies
 - Applications: Data mining, Semantic Indexing, Semi-automatic annotation, Text summarization

Future Work

- Semantic Annotation
 - Incorporate word sense disambiguation
 - Use word discrimination (IPF) with other scoring approaches
- Text Summarization
 - Multiple-document summarization using concepts
 - Improve redundancy removal
 - Incorporate text structure
 - Personalized Summaries

Publication Summary

- 2 Journal Papers (refereed)
- 2 Book Chapters (refereed)
- 5 Conference Papers (refereed), 1 submitted
- 1 Bulletin Paper, 1 DUC paper
- 5 posters, 1 received IST Dean's Award

Publications

2007:

- Reeve, L. H., & Han, H. (2007). CONANN: An Online Biomedical Concept Annotator. *Proceedings of the 2007 Data Integration in the Life Sciences conference (DILS'07)*, Philadelphia, PA USA.
- Lawrence Reeve and Hyoil Han (2007). *Online Biomedical Concept Annotation Using Word Coverage Filtering and Language Model Mapping*. Submitted.
- Lawrence Reeve, Hyoil Han and Ari D. Brooks (2007). *Biomedical Text Summarization Using Concept Chains*, International Journal of Data Mining and Bioinformatics. Refereed. Accepted.
- Lawrence Reeve, Hyoil Han and Ari D. Brooks. *The Use of Domain-Specific Concepts in Biomedical Text Summarization*, Journal of Information Processing and Management. Refereed. Accepted.
- Reeve, L. H., & Han, H. (2007). *A Term Frequency Distribution Approach for the DUC-2007 Update Task*. Proceedings of the 2007 Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007), Rochester, NY, USA.

Publications, cont'd

2006:

- Lawrence H. Reeve, Hyoil Han, Saya V. Nagori, Jonathan C. Yang, Tamara A. Schwimmer, and Ari D. Brooks (2006). *Concept Frequency Distribution in Biomedical Text Summarization*. Proceedings of the 15th Conference on Information and Knowledge Management. *Refereed* - 15% acceptance.
- Lawrence Reeve, Hyoil Han, and Ari D. Brooks (2006). *BioChain: Using Lexical Chaining Methods for Biomedical Text Summarization*. Proceedings of the 21st Annual ACM Symposium on Applied Computing, Bioinformatics track. *Refereed* - 32% acceptance.
- Lawrence Reeve and Hyoil Han (2006). *A Comparison of Semantic Annotation Systems for Text-based Web Documents*. *Web Semantics and Ontology*, David Taniar and J. Wenny Rahayu (Eds.), Idea Group Publishing. *Refereed*.

Publications, cont'd

2005:

- Lawrence Reeve and Hyoil Han (2005). *Survey of Semantic Annotation Platforms*. Proceedings of the 20th Annual ACM Symposium on Applied Computing, Web Technologies and Applications track. Presentation. *Refereed* - 37% acceptance.
- Lawrence Reeve and Hyoil Han (2005). *Semantic Annotation for Semantic Social Networks Using Community Resources*. AIS SIGSEMIS Bulletin, Vol. 2, Issue (3&4), pp: 52-56.
- Lawrence Reeve, Hyoil Han, and Chaomei Chen (2005). *Information Visualization and the Semantic Web*. Visualizing the Semantic Web, Vladimir Geroimenko and Chaomei Chen (Eds.), Springer. *Refereed*.

2004:

- Lawrence Reeve (2004). *Adapting the TileBar Interface for Visualizing Resource Usage*. Proceedings of the 30th International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. Presentation. *Refereed*.

References

- Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. *Proceedings of the AMIA Symposium 2001*, 17-21.
- Aronson, A. R. (1996). *MetaMap: Mapping text to the UMLS metathesaurus*. Unpublished manuscript.
- Barzilay, R., & Elhadad, M. (1997). Using lexical chains for text summarization. In Proceedings of the Intelligent Scalable Text Summarization Workshop (ISTS'97), ACL, Madrid, Spain. 10-18. Aronson, A. R. (2001). Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. *Proceedings of the AMIA Symposium 2001*, 17-21.
- Brooks, A. D., & Sulimanoff, I. (2002). Evidence-based oncology project. *Surgical oncology clinics of north america* (pp. 3-10)
- Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in Bioinformatics*, 6(1), 57-71.
- Dalianis, H. (2000). *SweSum - A Text Summarizer for Swedish* No. TRITA-NA-P0015). Stockholm, Sweden: NADA, KTH.
- Denny, J. C., Smithers, J. D., Miller, R. A., & Spickard, A. (2003). "Understanding" medical school curriculum content using KnowledgeMap. *Journal of the American Medical Informatics Association*, 10(4), 351-362.
- Denny, J. C., Irani, P. R., Wehbe, F. H., Smithers, J. D., & Spickard, A.,3rd. (2003). The KnowledgeMap project: Development of a concept-based medical school curriculum database. *Proceedings of the Annual AMIA Symposium*, , 195-199.
- Edmundson, H. P. (1999). New methods in automatic extracting. In I. Mani, & M. T. Maybury (Eds.), (pp. 23-42). Cambridge, MA: MIT Press.
- Galley, M., & McKeown, K. (2003). Improving word sense disambiguation in lexical chaining. *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, Acapulco,Mexico. 1486-1488.
- Goldstein, J., Kantrowitz, M., Mittal, V., & Carbonell, J. (1999). Summarizing text documents: sentence selection and evaluation metrics. *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, Berkeley, California, United States. 121-128

References, cont'd

- Hersh, W. R., & Greenes, R. A. (1990). SAPHIRE--an information retrieval system featuring concept matching, automatic indexing, probabilistic retrieval, and hierarchical relationships. *Computers and Biomedical Research, an International Journal*, 23(5), 410-425.
- Kupiec, J., Pedersen, J., & Chen, F. (1999). A trainable document summarizer. In I. Mani, & M. T. Maybury (Eds.), (pp. 55-60). Cambridge, MA: MIT Press.
- Lin, C. (2004). Looking for a few good metrics: Automatic summarization evaluation - how many samples are enough? *Proceedings of the NTCIR Workshop 4*, Tokyo, Japan.
- Lin, C. Y., & Och, F. J. (2004a). Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain. 605-612.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.
- Mani, I., & Maybury, M. T. (1999). In Mani I., Maybury M. T. (Eds.), *Advances in automatic text summarization*. Cambridge, MA: MIT Press.
- Mani, I., & Bloedorn, E. (1998). Machine Learning of Generic and User-Focused Summarization. *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/ Innovative applications of artificial intelligence*, Madison, Wisconsin, United States. 820-826.
- Morris, J., & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1), 21-43.

References , cont'd

- Nadkarni, P. M. (1997). Concept locator: A client-server application for retrieval of UMLS metathesaurus concepts through complex boolean query. *Computers and Biomedical Research, an International Journal*, 30(4), 323-336.
- National Institute of Health. (2005). *Identification of Important Text in Full Text Articles Using Summarization*.
- Nenkova, A., & Vanderwende, L. (2005). *The impact of frequency on summarization* No. MSR-TR-2005-101). Redmond, Washington: Microsoft Research.
- Pollock, J. J., & Zamora, A. (1975). Automatic abstracting research at chemical abstracts service. *Journal of Chemical Information and Computer Sciences*, 15(4), 226-232.
- Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Celebi, A., Drabek, E., et al. (2004). MEAD - a platform for multidocument multilingual text summarization. *Proceedings of Language Resources and Evaluation 2004 (LREC 2004)*, Lisbon, Portugal.
- Rose, P. G., Bundy, B. N., Watkins, E. B., Thigpen, J. T., Deppe, G., Maiman, M. A., et al. (1999). Concurrent Cisplatin-Based Radiotherapy and Chemotherapy for Locally Advanced Cervical Cancer. *New England Journal of Medicine*, 340(15), 1144-1153.
- Rotem, N. (2003). *Open Text Summarizer (OTS)*. Retrieved July 3, 2006, 2006, from <http://libots.sourceforge.net>
- Silber, G. H., & McCoy, K. F. (2002). Efficiently computed lexical chains as an intermediate representation for automatic text summarization. *Computational Linguistics*, 28(4), 487-496.
- Sparck Jones, K. (1999). Automatic summarizing: Factors and directions. In I. Mani, & M. T. Maybury (Eds.), *Advances in automatic text summarization* (pp. 2-12). Cambridge, MA: MIT Press.
- Srinivasan, S., Rindflesch, T. C., Hole, W. T., Aronson, A. R., & Mork, J. G. (2002). Finding UMLS metathesaurus concepts in MEDLINE. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, , 727-731.
- Strzalkowski, T., Stein, G., & Wang, J. B, wise. A robust practical text summarizer. 1999. *Advances in Automatic Text Summarization*, , 137–154.

References , cont'd

- Teufel, S., & Moens, M. (1999). Argumentative classification of extracted sentences as a first step towards flexible abstracting. In I. Mani, & M. T. Maybury (Eds.), (pp. 155-171). Cambridge, MA: MIT Press.
- The Lemur Project. (2006). *Lemur Language Modeling Toolkit*. Retrieved July 3, 2006, 2006, from <http://www.lemurproject.org/>
- United States National Library of Medicine. (2005a). *ClinicalTrials.gov*.<http://www.clinicaltrials.gov/>
- United States National Library of Medicine. (2005d). *Unified medical language system (UMLS)*.<http://www.nlm.nih.gov/research/umls/>
- United States National Library of Medicine. (2006a). *PubMed*.<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>
- Vanderwende, L., & Suzuki, H. (2005). Frequency-based summarizer and a language modeling extension. *Proceedings of the MultiLingual Summarization 2005 Workshop*,
- Wollersheim, D., Rahayu, W., & Reeve, J. (2002). Evaluation of index term discovery in medical reference text. Paper presented at the *Proceedings of the International Conference on Information Technology and Applications*, Bathurst, NSW, Australia.
- Zieman, Y. L., & Bleich, H. L. (1997). Conceptual mapping of user's queries to medical subject headings. *Proceedings : A Conference of the American Medical Informatics Association / AMIA Annual Fall Symposium. AMIA Fall Symposium*, , 519-522.
- Zou, Q., Chu, W. W., Morioka, C., Leazer, G. H., & Kangarloo, H. (2003). IndexFinder: A method of extracting key concepts from clinical texts for indexing. *Proceedings of the AMIA Annual Symposium*, 763-767.