A Short Survey of Discourse Representation Models

Tudor Groza, Siegfried Handschuh, Tim Clark, Simon Buckingham Shum and Anita de Waard

Semantic Web Applications in Scientific Discourse
Workshop @ ISWC 2009

26 October 2009
Chantilly, Virginia, US

Outline

- **Introduction**
- **Analysis features**
- **Models**
- **Analysis overview**
- **Conclusion**

Outline

- **Introduction**
- Analysis features
- Models
- Analysis overview
- Conclusion

Introduction

- **Dissemination – communication process**
- **Externalization**
  - Implicit – materialized as publications
  - Explicit – required for machine interpretation
- **Discourse structuring and analysis**
  - Discourse representation models
  - Computational linguistic approaches – automatic extraction of epistemic items

Introduction (cont.)

- **Our focus: Discourse representation models**
  - Succinct overview
  - Brief comparative analysis

- **Tentative goal: an unified discourse representation model**

Outline

- Introduction
- **Analysis features**
- Models
- Analysis overview
- Conclusion

Analysis features

- **Coarse-grained rhetorical structure**
- **Fine-grained rhetorical structure**
- **Relations** – types of relations used
- **Polarity** – explicit positive vs. negative
- **Weights** – explicit numeric weight of relations
- **Provenance** – localization in text
- **Shallow metadata support**
- **Domain knowledge**
- **Purpose** – intended use of the model
- **Evaluation and uptake**

Outline

- Introduction
- Analysis features
- **Models**
- Analysis overview
- Conclusion

Models

- Harmsze's model
- The Scholarly Ontologies project
- De Waard's model
- The SWAN Ontology
- The SALT Framework

Harmsze's model

- **Purpose**: presentation of scientific information in electronic articles
- **Coarse-grained structure**
  - Modules and elementary modules
  - E.g.: Meta-information, Positioning, Methods, Results, Interpretation, Outcome
- **Relations**
  - Organizational links
  - Scientific discourse links
    - Communicative function – *elucidation*, *argumentation*, *clarification*
    - Content relations – *elaboration*, *aggregation*

The Scholarly Ontologies Project

- **Purpose**: General structuring of coherence and argumentation
- **Fine-grained structure**
  - Atomic nodes – short pieces of text
  - Claims – connected nodes
- **Relations**
  - Cognitive Coherent Relations – Sanders et al.
  - Explicit polarity and weights – *proves (+1)* vs. *refutes (-1)*
  - Types: causal, problem related, similarity, general, supports/challenges, taxonomic

De Waard's model

- **Purpose**: Modularization of scientific publications
- **Coarse-grained structure**
  - **A**nnotation – **B**ackground – **C**ontribution – **D**iscussion – **E**ntities
- **Relations**
  - Argumentative
  - Explicit polarity
  - E.g.: *proves* vs. *refutes*; *agrees* vs. *disagrees*

The SWAN Ontology

- **Purpose**: Creation of knowledge bases
  - Initially in the context of the Alzheimer Disease Research
- **General structure**
  - 6 main elements: people, bibliographic records, life science entities, tags, versions, discourse elements
- **Fine-grained structure**
  - Discourse Element, Research Statement, Research Question, Structure Comment
- **Relations**
  - Argumentative
  - E.g.: *consistentWith*, *inconsistentWith*, *discusses*

The SALT Framework

- **Purpose**: Structuring of rhetoric and argumentation in scientific publications
- **General structure**
  - 3 layers
- **Coarse-grained structure**
  - Rhetorical blocks: Introduction, Conclusion, …
- **Fine-grained structure**
  - Rhetorical elements: Claims, Supports, …
- **Relations**
  - Rhetorical relations (Rhetorical Structure of Text – Mann et al.): *Antithesis*, *Consequence*, …
  - Argumentative relations

Outline

- Introduction
- Analysis features
- Models
- **Analysis overview**
- Conclusion

Analysis overview

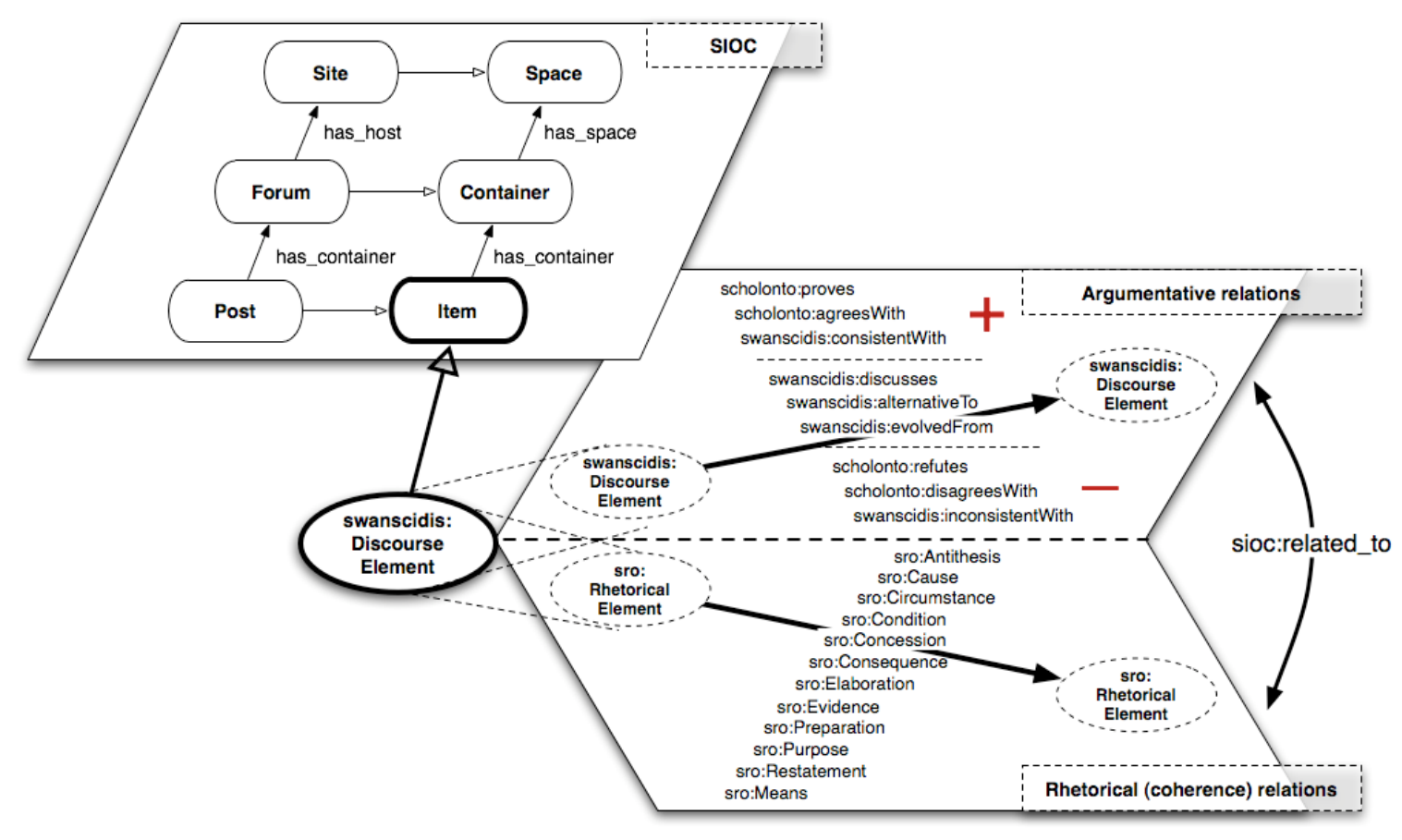
Feature/ Approach	Coarse-grained rhetorical structure	Fine-grained rhetorical structure	Relations	Polarity	Weights	Provenance	Shallow Metadata Support	Domain Knowledge	Purpose	Evaluation and uptake
Harmsze	Modules	Elementary module	Structuring, organisational and discourse	Implicit (within relations)	No	No	Yes	Open	Presentation of scientific information in electronic articles	Preliminary evaluation
ScholOnto	No	Node, Claim	Cognitive Coherent	Explicit (+ / -)	Explicit (1, 2)	Yes (duplicates)	No	Open	General structuring of coherence and argumentation	Evaluated and widely used
De Waard	Rhetorical Blocks	Rhetorical element	Argumentative	Explicit (within the pairs of relations)	No	No	Yes	Open	Modularization of scientific publications	N/A
SWAN	No	Discourse element, Research statement	Argumentative and Cognitive Coherent	Implicit (within relations)	No	Yes (duplicates)	Yes	Yes (Gene, Protein)	Creation of a knowledge base	Evaluated and widely used
SALT	Rhetorical Blocks	Rhetorical element (Nucleus, Satellite, Claim, Support)	Rhetorical and argumentative	Implicit (within relations)	No	Yes (pointers)	Yes	Open	Structuring of the rhetoric and argumentation in scientific publications	Evaluated and infrequenly used

Towards an unified discourse representation model

- **Proper balance of currently existing features**
- **Emphasis on practicality for uptake maximization**
- **General structure**
  - Layered – e.g. SWAN, SALT
- **Coarse-grained structure**
  - Rhetorical blocks – e.g. ABCDE, SALT
- **Fine-grained structure**
  - Discourse elements
- **Relations**
  - 2 layers
  - Argumentative
  - Rhetorical relations

Abstract layering view

Concrete (Web-oriented) Example



Outline

- Introduction
- Analysis features
- Models
- Analysis overview
- **Conclusion**

Conclusion

- **Succinct overview of current discourse representation models**
- **Brief comparative analysis**

- **Next steps:** *... open for discussion*

Thank you!