

422 big picture: Where are we With reading papers?

Hybrid: Det + Sto

Prob CFG
Prob Relational Models
Markov Logics

Deterministic

Stochastic

<p>Logics First Order Logics</p> <p>Ontologies</p> <ul style="list-style-type: none"> Full Resolution SAT 	<p>Belief Nets</p> <p>Approx. : Gibbs</p> <p>Markov Chains and HMMs</p> <p>Forward, Viterbi....</p> <p>Approx. : Particle Filtering</p> <p>Undirected Graphical Models Markov Networks</p> <p>Conditional Random Fields</p>
<p>Planning</p>	<p>Markov Decision Processes and Partially Observable MDP</p> <ul style="list-style-type: none"> Value Iteration Approx. Inference <p>Reinforcement Learning</p>

Query

Planning

Applications of AI

Representation

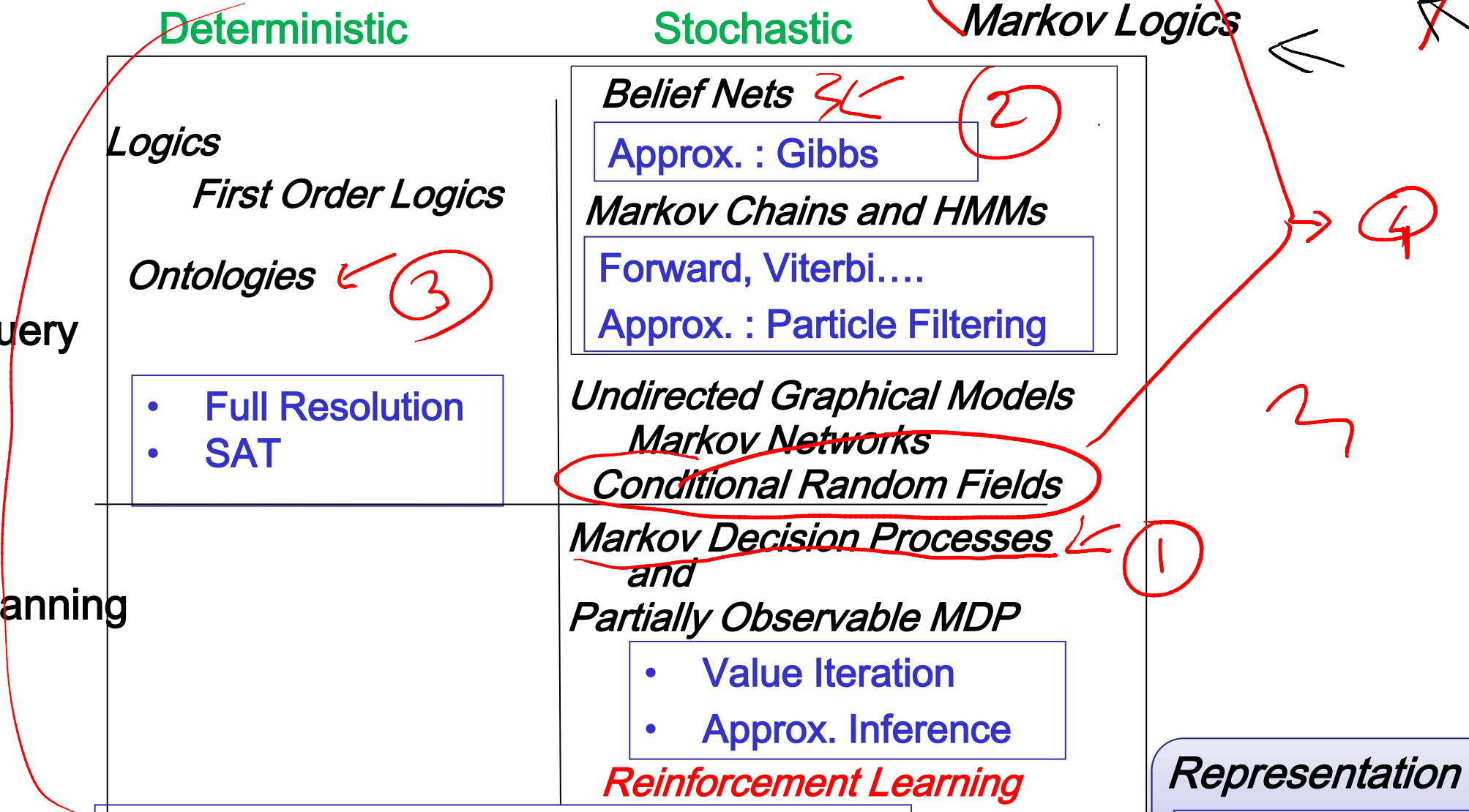
Reasoning
Technique

3

2

4

1



Extracting Knowledge from Evaluative Text

Giuseppe Carenini, Raymond T. Ng, Ed Zwart

Computer Science Dept.

University of British Columbia

Vancouver, CANADA

Motivation and Focus

- Large amounts of info expressed in **text form** is constantly produced
 - News, Reports, Reviews, Blogs, Emails....
- Pressing need to extract and summarize key/strategic info
- Considerable work but limited factual info

Our Focus: evaluative text about single entity (good vs. bad, right vs. wrong)

- Customer reviews
- Travel logs
- Job candidate evaluations..... etc.

KCAP from evaluative text (single entity)

- Extract relevant knowledge

A. What features of the entity are evaluated in the reviews? [Hu, Liu AAI '04] [Popescu Etzioni HLT '05]

B. For each feature:

1. what is the polarity of the evaluation? (good vs. bad)

[Hu, Liu KDD '04]

2. what is the strength of the evaluation? (rather good vs. extremely good) [Wilson et al. AAI '04]

- Summarize and present extracted knowledge to user

Outline

- Feature Extraction: limitations of previous work and our solution
- Evaluation of our approach
- Benefits in term of KCAP
- Conclusion and Demo of Future work 😊

Feature Extraction: sample from corpus

[Hu&Liu 2004]

..... the canon computer software used to download , sort , . . . is very nice and very easy to use. the only two minor issues i have with the camera are the lens cap (it is not very snug and can come off too easily)

the menus are easy to navigate and the buttons are easy to use. it is a fantastic camera and well worth the price .

Feature Extraction: sample from corpus

[Hu&Liu 2004]

..... the canon **computer software** used to download , sort , . . . is very nice and very easy to use. the only two minor issues i have with the camera are the **lens cap** (it is not very snug and can come off too easily)

the **menus** are easy to navigate and the **buttons** are easy to use. it is a fantastic **camera** and well worth the price .

Limitations of previous approach

Key problems with extracted features (for KCAP):

- May be too many and redundant (often > 100)
- Flat, unstructured list (lack hierarchical organization)
- May be expressed in an unfamiliar terminology (for target user)

- spot metering
- metering option
- remote control
- battery
- night mode
- light automatic correction

- battery life
- remote
- battery charging system
- low light focus
- ...

Battery

Lighting

Example Ideal Mapping

UDFs

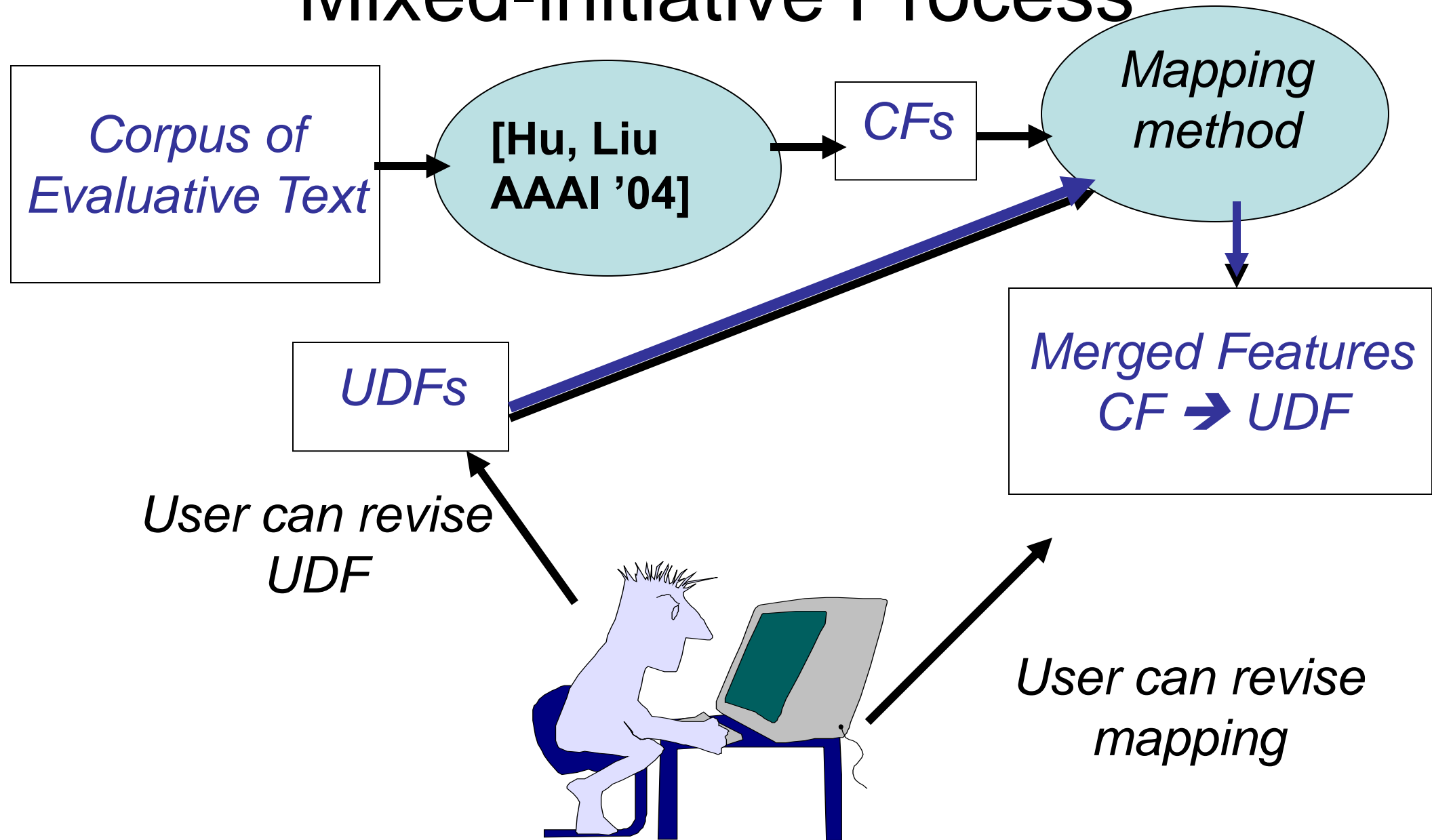
CFs

1. Canon G3 PS Digital Camera [*canon, canon PS g3, digital camera, camera,...*]
 - 1.1 User Interface [*button, menus, lever*]
 - 1.2 Convenience []
 - Battery [*battery life, battery charging system, battery*]
 - Self Timer []
 - Burst Mode [*speed, mode*]
 - Rapid Fire Shot [*speed*]
 - Delay Between Shots [*unresponsiveness, delay, speed, lag time, lag*]
 -
2. Not Placed [*manual, function, quality, strap, service, shoot, learning curve,...*]

Our Solution

- Map extracted features (**Crude Features (CF)**) in a hierarchy of product features at different levels of abstraction. Two alternatives:
 - Learn the hierarchy
 - – Have the user provide a hierarchy of **User Defined Features (UDF)**
- Such a mapping will:
 - Eliminate redundancy (CFs with same referent mapped in the same UDF)
 - Provide conceptual structure
 - Increase user familiarity with CFs

Mixed-initiative Process



Our Mapping Method

- Map each CF in the “most similar” UDFs
- CFs and UDFs are terms (i.e., sequences of 1 or more words)

So need a measure of term similarity

- **Our approach** to term similarity: combine similarity between constituent words
- **So need a measure of word similarity and a function to combine similarities**

Word Similarity Metrics *w_m*

- **String Matching:** baseline
 - **WordNet Synset Matching:** ↙
1 if the two words appear in the same synset....
e.g., (photograph, photo, exposure, picture)
 - **Wordnet Distance Matching:** a set of measures ↘
that compute the semantic distance between the
synsets of the two words
- [Patwardan, Pedersen 2003] Cpan module

Term similarity: Combine word scores

$$cf = \{v_1, \dots, v_n\}; \quad udf = \{w_1, \dots, w_m\}$$

- **MAX:** terms' score is the maximum score of comparing all possible word pairs

$$\max_{i,j} \{wm(v_i, w_j)\}$$

- **AVG:** terms' score is the average of the max of all i with j , and vice versa (to avoid a high score of one word dominate the whole term's score)

$$\frac{\sum_{i=1}^n \max_j \{wm(v_i, w_j)\}}{n} + \frac{\sum_{j=1}^m \max_i \{wm(v_i, w_j)\}}{m} / 2$$

Mapping Algorithm

Algorithm

- Each CF is mapped to the UDF with which it receives the greatest similarity score
- In case of tie scores CF is mapped more than once
- **But** mapping occurs only if score greater than a given threshold

Threshold

- For **string** and **synset** matching the threshold was set to 0.
- For **Wordnet distance** similarity measures was set by varying a parameter θ

Outline

- Feature Extraction: limitations of previous work and our solution
- Evaluation of our approach
- Benefits in term of KCAP
- Conclusion and Demo of Future work 😊

Results DigCam for AVG

		Mapping quality measures	
		p_dist	redun
Word Similarity metrics	str_match	.38	.23
	syn_score	.39	.27
Wordnet distance (lin)	θ		
	-0.2	.35	.29
	-0.4	.39	.33
	-0.6	.43	.40

↓ 😊 ↑ 😊

Results DVD for AVG

Mapping quality measures

Word Similarity metrics

	p_dist	redun
str_match	.27	.21
syn_score	.28	.25
θ		
-0.2	.40	.38
-0.4	.47	.45
-0.6	.55	.52

Wordnet distance (lin)



Outline

- Feature Extraction: limitations of previous work and our solution
- Evaluation of our approach
- Benefits in term of KCAP
- Conclusion and Demo of Future work 😊

Benefits in term of KCAP

Key questions for manufactures and potential customers

- what product features are more frequently mentioned?
- how do customers evaluate those features?
- do they agree?

CFs only

Table 3: *CF* Frequency Statistics

Crude Feature	Total	Pos	Neg
camera	57	55	2
picture	15	13	2
viewfinder	12	1	11
...			
lcd	3	3	0
...			
image quality	1	1	0
image	1	1	0
display	1	1	0
shot	1	1	0

Benefits in term of KCAP

- ... answer the same questions
- different levels of abstraction
- less redundancy
- more familiar terms

CF UDF mapping

Table 5: Informative Mapping Results

UDF	Total	Pos	Neg
Editing/Viewing	17	6	11
LCD Display	4	4	0
Viewfinder	12	1	11

Table 4: Reduced Redundancy

UDF	Total	Pos	Neg
Image	51	46	5
Image Type	2	1	1
TIFF	1	0	1
...			
Resolution	3	2	1
Effective Pixels	2	2	0
...			

**image
picture
shot**

Conclusions

- **Novel approach to feature extraction step in KCAP from evaluative text**
- Mixed-initiative mapping of flat list of extracted CF into a UDF hierarchy
- Term similarity metrics
- **Evaluation of these metrics** on two corpora of customer reviews : reasonable accuracy, substantial reduction in redundancy
- **Beneficial in term of captured knowledge**

Future Work

- Improve mapping method
 - Try other term similarity measures (corpus based)
 - Inject more sophisticated NLP (e.g., weight scoring considering headword of a term)
- Develop interface to support user revision of the mapping and of the UDF hierarchy
- Summarize and present extracted knowledge to user Combine text and graphics....
Adapt techniques for generating evaluative text

Questions 2015-2

- Is WordNet the best online lexical database?!? 😊
- Who is the user?
- UDFs / CFs / Gold Standard
- Unplaced CFs
- CF extraction and polarity (how many methods?)
- Constructing large UDF
- Different Languages ←
- Threshold
- Future
 - Microsoft Research took this over in 2007-8 ↙
 - Interactive Multimedia Summarization (Visualization) ↙
 - Lexical Similarity vs. corpus-based ←
 - Automatically create UDFs: Extract Hierarchy from the reviews/ from existing ontologies - Speech input... Sarcasm



https://www.google.com/patents/US8799773

Apps UBC Research Funding +... Bookmarks my-tabs-Aug29 The AI Behind Wats... UBC Policy and Decision-... Societa carenini A Statistical NLG Fra... NAACL H



Search bar with magnifying glass icon

Patents

Application

Grant

Find prior art

Discuss this patent

View P

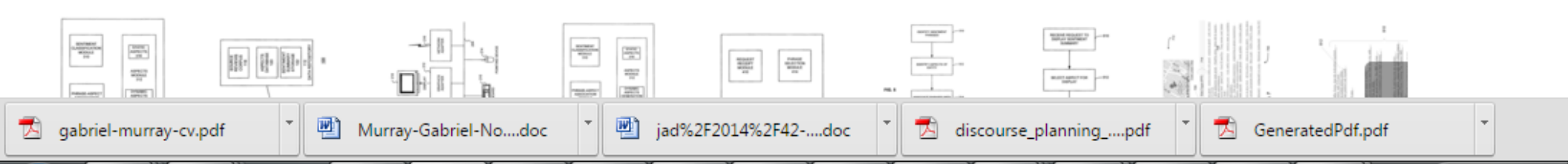
Aspect-based sentiment summarization US 8799773 B2

ABSTRACT

Phrases in the reviews that express sentiment about a particular aspect are identified. Reviewable aspects of the entity are also identified. The reviewable aspects include static aspects that are specific to particular types of entities and dynamic aspects that are extracted from the reviews of a specific entity instance. The sentiment phrases are associated with the reviewable aspects to which the phrases pertain. The sentiment expressed by the phrases associated with each aspect is summarized, thereby producing a summary of sentiment associated with each reviewable aspect of the entity. The summarized sentiment and associated phrases can be stored and displayed to a user as a summary description of the entity.

Publication number	US8799773 B2
Publication type	Grant
Application number	US 12/051,798
Publication date	Aug 5, 2014
Filing date	Mar 19, 2008
Priority date ?	Jan 25, 2008
Also published as	CA2711665A1, 5 More »
Inventors	George Reis, Sasha Blair-Goldensohn, Ryan T. McDonald
Original Assignee	Google Inc.
Export Citation	BiBTeX, EndNote, RefMan
Patent Citations (44), Non-Patent Citations (57), Referenced by (4), Classifications (13), Legal Events (1)	
External Links:	USPTO, USPTO Assignment, Espacenet

IMAGES (9)



gabriel-murray-cv.pdf Murray-Gabriel-No....doc jad%2F2014%2F42-....doc discourse_planning_....pdf GeneratedPdf.pdf

Data and Gold Standard

Two products: Digital Camera and DVD

- **CFs** from Hu&Liu annotated corpora: 101 CFs for digital camera, 116 for DVD
- **UDFs** developed by domain experts: 86 UDFs for digital camera, 38 for DVD

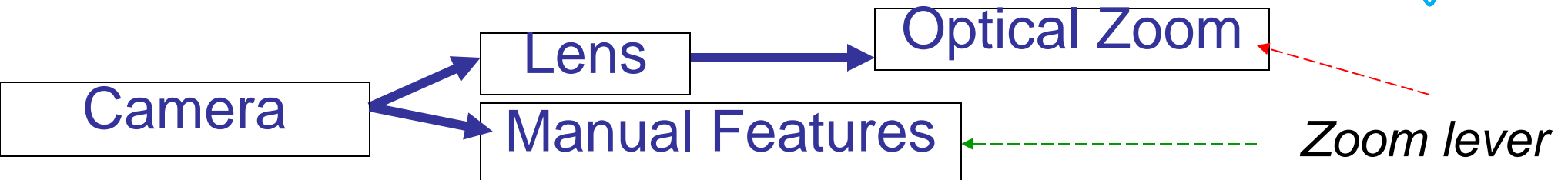
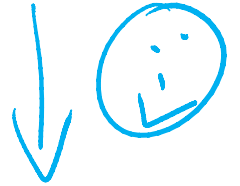
Gold Standard Development:

- We manually developed initial mappings
- User study: we asked 7 subjects to fix our mappings with some random errors
- Based on their input a final version was created

Measures of mapping quality

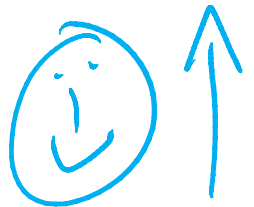
- (Graphical “distance” from correct placement)

$$placement_distance(cf_i) = avg(edgeCount(cf_i))$$



- (Fraction of redundant CF's)

$$redun_reduc = \frac{|placedCF| - |nonEmptyUDF|}{|CF|}$$



Can be maximized by placing all CFs in one UDF but...

redun_reduc in Gold Stand. DCam = .45 ; DVD=.43

$$\text{sim}_{\text{path}}(c_1, c_2) = -\log \text{pathlen}(c_1, c_2)$$

$$\text{sim}_{\text{Resnik}}(c_1, c_2) = -\log P(\text{LCS}(c_1, c_2))$$

sim_{res}

$$\text{sim}_{\text{Lin}}(c_1, c_2) = \frac{2 \times \log P(\text{LCS}(c_1, c_2))}{\log P(c_1) + \log P(c_2)}$$

$$\text{sim}_{\text{jc}}(c_1, c_2) = \frac{1}{2 \times \log P(\text{LCS}(c_1, c_2)) - (\log P(c_1) + \log P(c_2))}$$

$$\text{sim}_{\text{eLesk}}(c_1, c_2) = \sum_{r, q \in \text{RELS}} \text{overlap}(\text{gloss}(r(c_1)), \text{gloss}(q(c_2)))$$

Talk Summary

*Corpus of
Evaluative
Documents*

[KCAP '05]
Extract evaluative info:
feature hierarchy annotated
with evaluations

SEA (NLG abstractor)
MEAD* (extractor)

**Generate
NL Summary**

[EACL '06]

[INLG '08]

**Allow access to
original sources:**
•Text footnotes
•Treemap zooming

**Present
As Treemaps**

**Present
NL Summary**

Multimedia Summary

Interactive

[IUI '06]

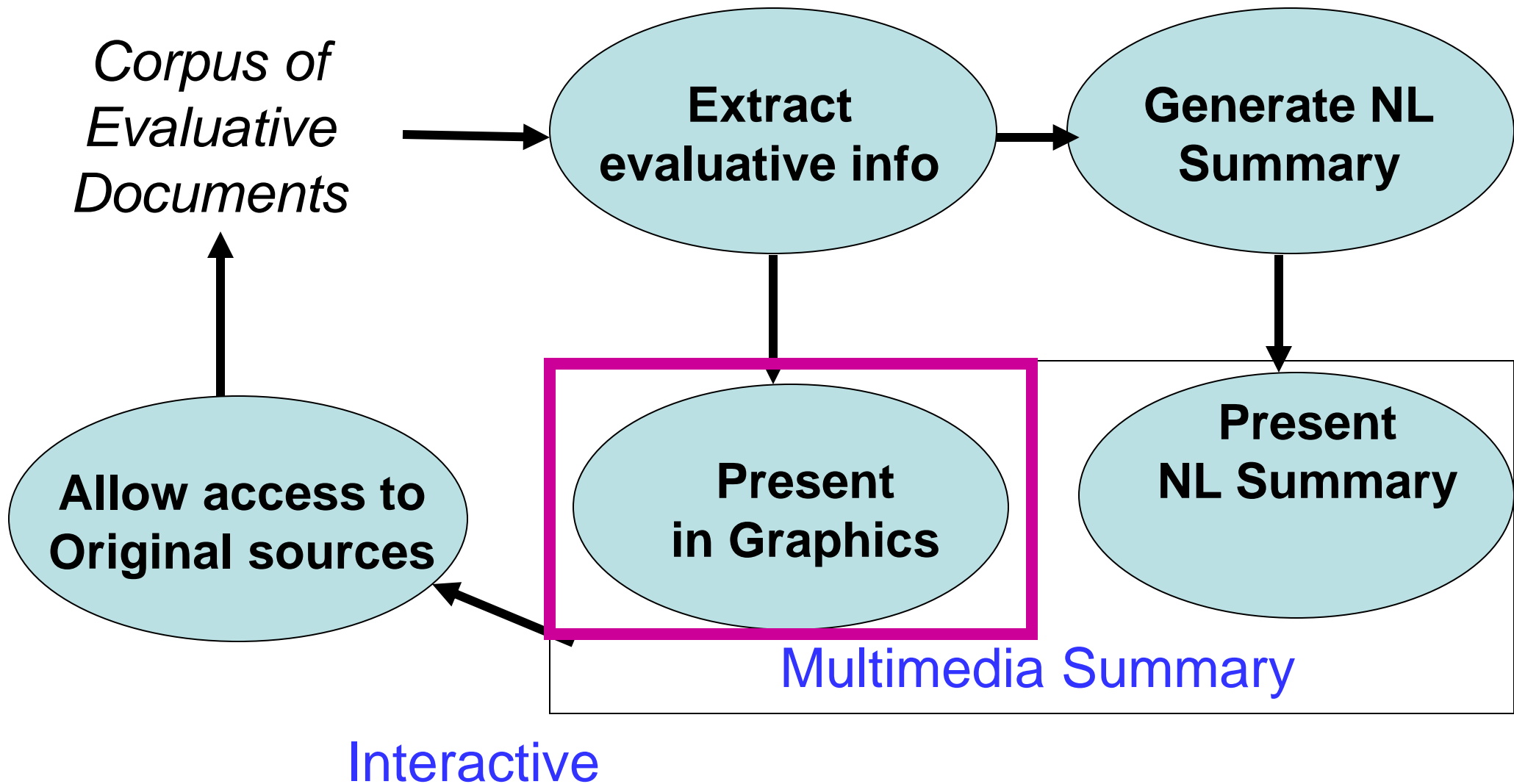
[IUI '09]

- G. Carenini , J. Cheung , A. Pauls. [Multi-Document Summarization of Evaluative text](#), **Computational Intelligence, 2012**
- Carenini G. and Rizoli L., A Multimedia Interface for Facilitating Comparisons of Opinions. . In: *Proceedings of the 13th International Conference on Intelligent User Interfaces, (IUI 2009)*, Sanibel Island, Sydney, Florida, USA, 2009 [[pdf](#)]
- Carenini G, Cheung J., Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality. *International Conference on Natural Language Generation. (INLG 2008)*, Salt Fork, Ohio, USA, June 12-14, 2008 [[pdf](#)]

Carenini G., Ng R., Pauls A. MultiDocument Summarization of Evaluative Text, *Proceedings of the 11th European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, Italy, April, 2006. [[pdf](#)]

Carenini G., Ng R., Pauls A. Interactive Multimedia Summaries of Evaluative Text, *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2006)*, Sydney, Australia, Gen29-Feb1, 2006. [[pdf](#)]

Multimedia Interactive Approach



Extracted evaluative info after mapping

- Merged Features hierarchy annotated with all the evaluations each feature received in the corpus

Canon G3 PS Digital Camera [-1,-1,+1,+2,+2,+3,+3,+3]

1. User Interface [+2]

– Button [+1]

– Menus [+2,+2,+2,+3+3]

– Lever []

2. Convenience []

– Battery []

• Battery life [-1,-1,-2]

• Battery charging system []

–

3.

PS_i is the set of polarity/strength evaluations for feature f_i

Conveying extracted info with graphics

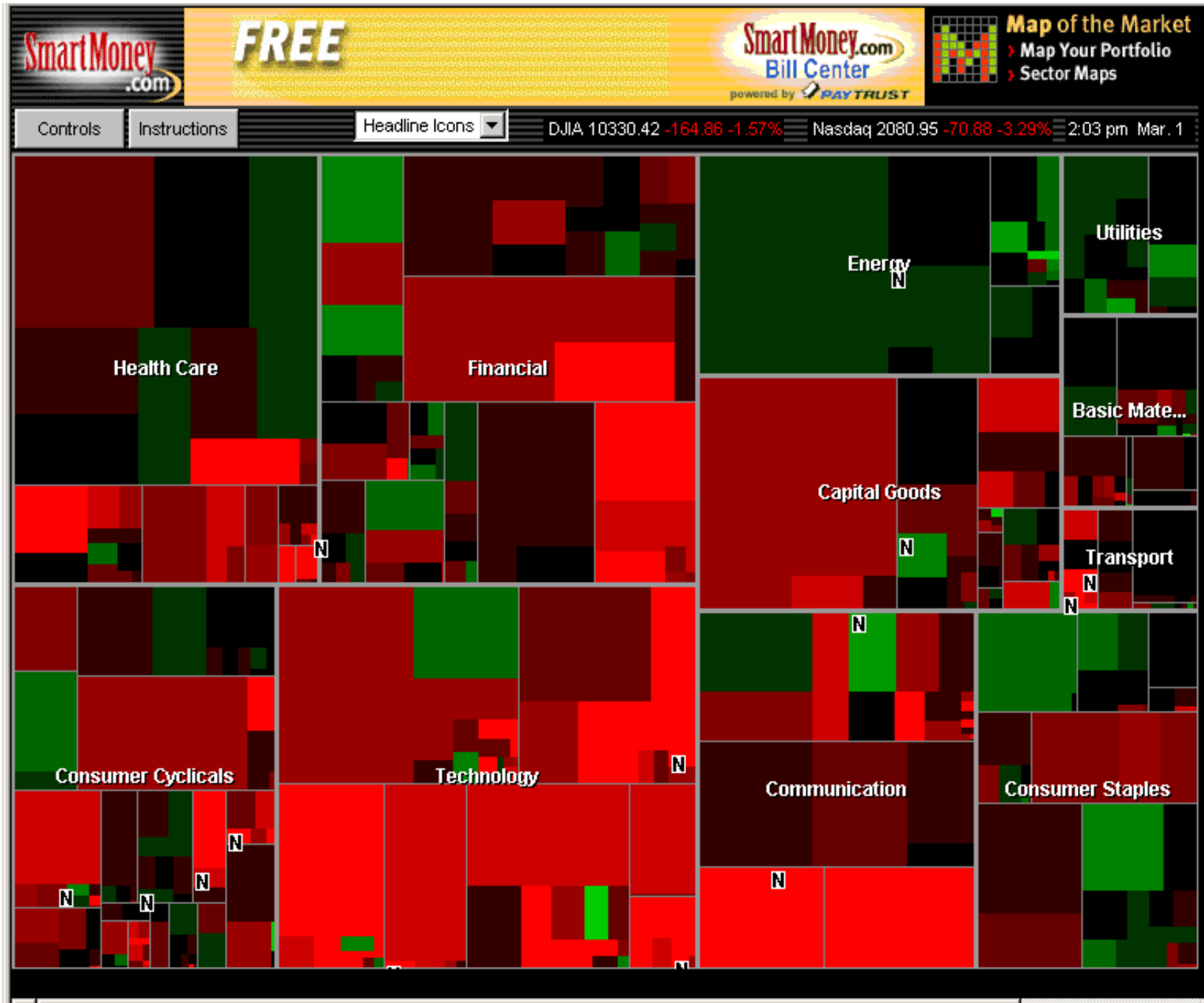
Visualization should convey:

- Hierarchical organization of the features
- For each feature
 - # of evaluations
 - polarity/strength of the evaluations (average?)

Treemaps: space-filling technique for visualizing hierarchical information structures

- Each node in the hierarchy is represented as a rectangle
- Descendants of a node are represented as nested rectangles
- Rectangle *size* and *colour* can express information about the node

Treemap: stock market



One possible Treemap

- Each product feature is represented as a rectangle
- The hierarchy is represented by nesting
- Rectangle *size* expresses # of evaluations
- Rectangle *colour* expresses avg polarity/strength of evaluations (black for neutral, the more **positive/negative** the more green/red)

Apex DVD Player

Disk Format

CD

DVD Audio

ormats

DVD-

Disc Formats self

Extra Features

Extra Features self

Universal Remote Control

General Information Apex AD2600 Progressive Scan DVD Player self

User Interface

Surround Sound Sup

Surround Sound

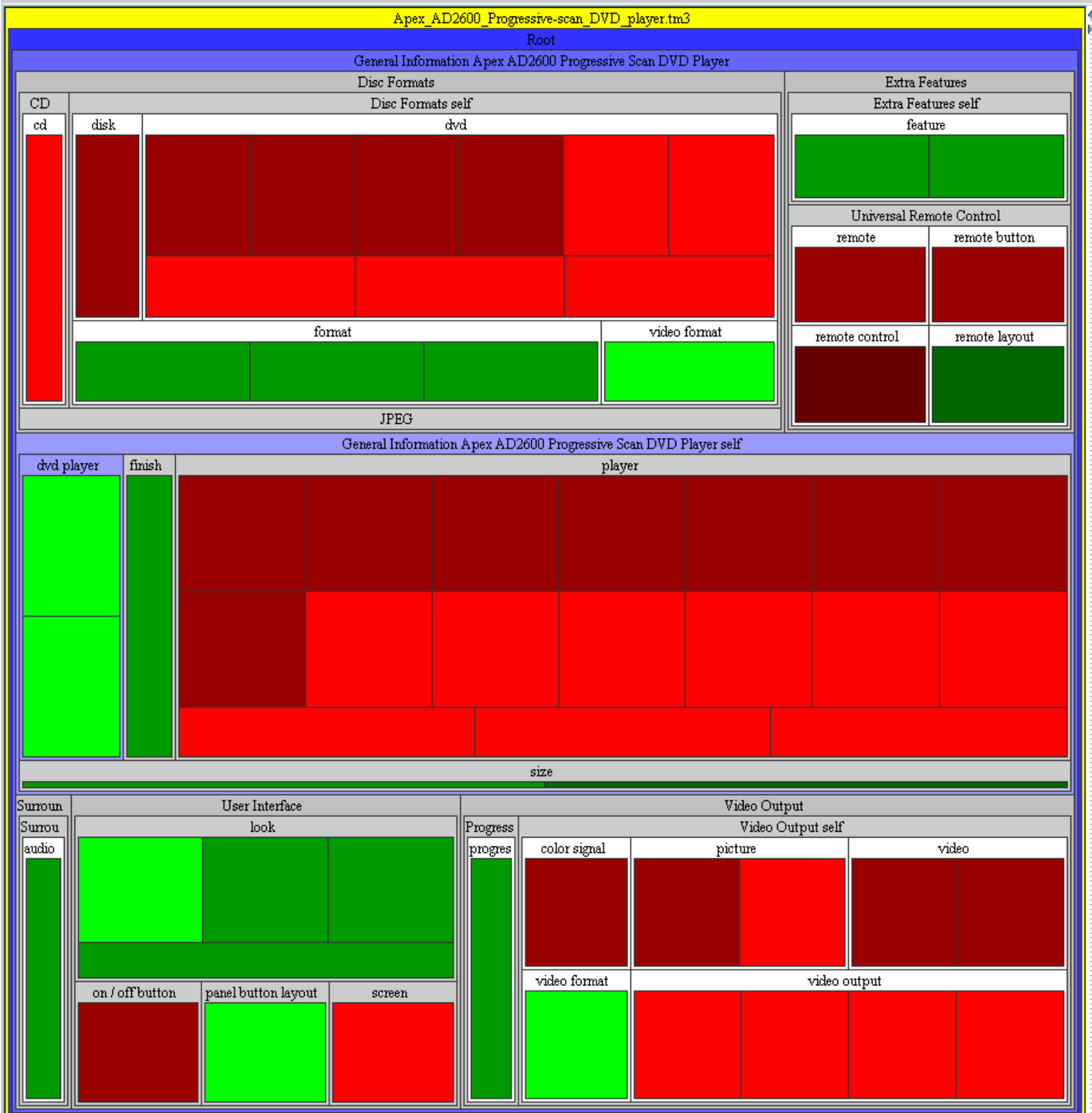
Progr S

Video Output

Video Output self

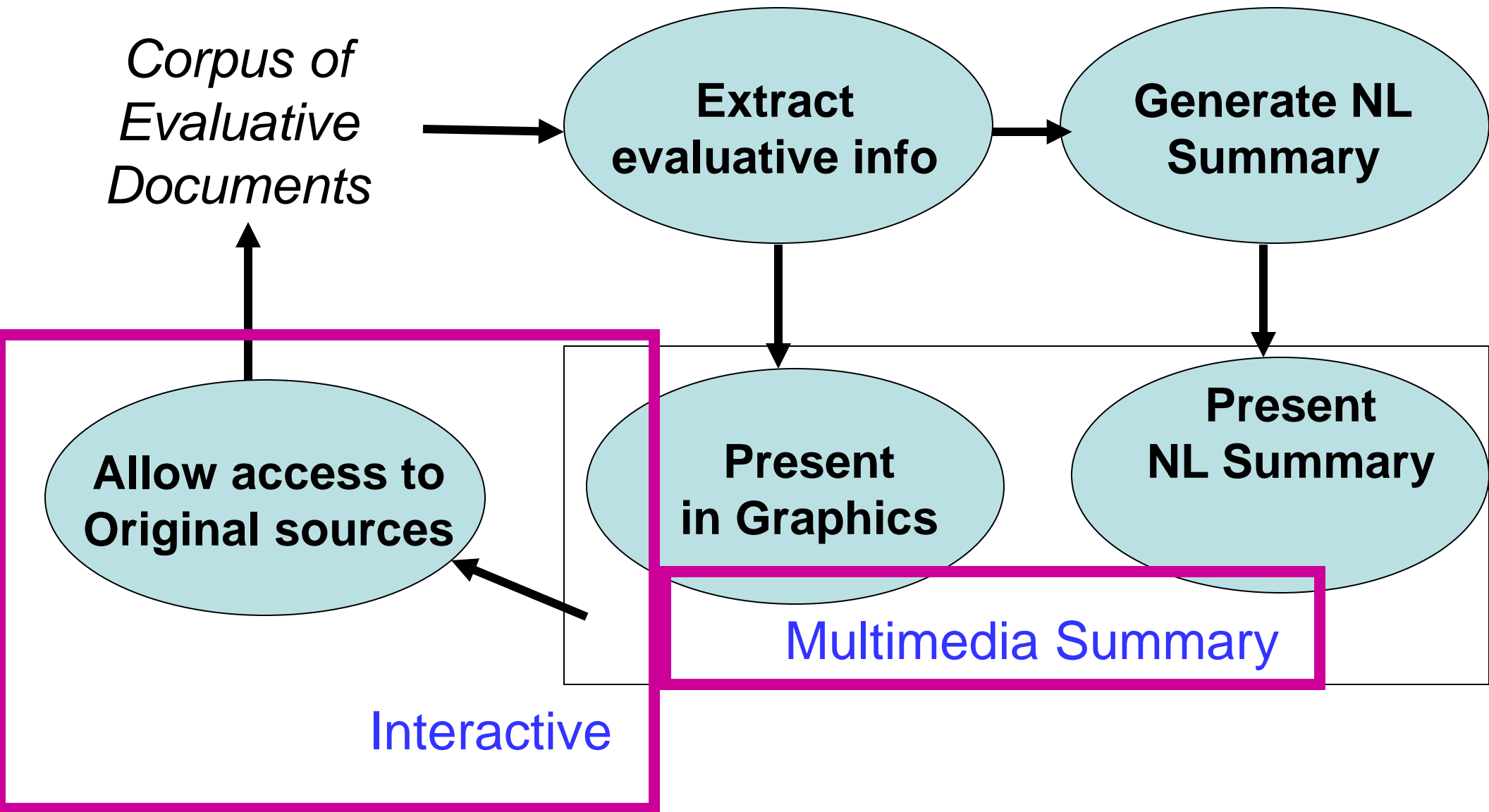
Another possible Treemap

- Each evaluation is represented as a rectangle
- The hierarchy is represented by nesting
- Rectangle *colour* expresses polarity/strength of the evaluation (black for neutral, the more **positive**/**negative** the more green/red)
- Note: More effective in conveying controversiality





Multimedia Interactive Approach

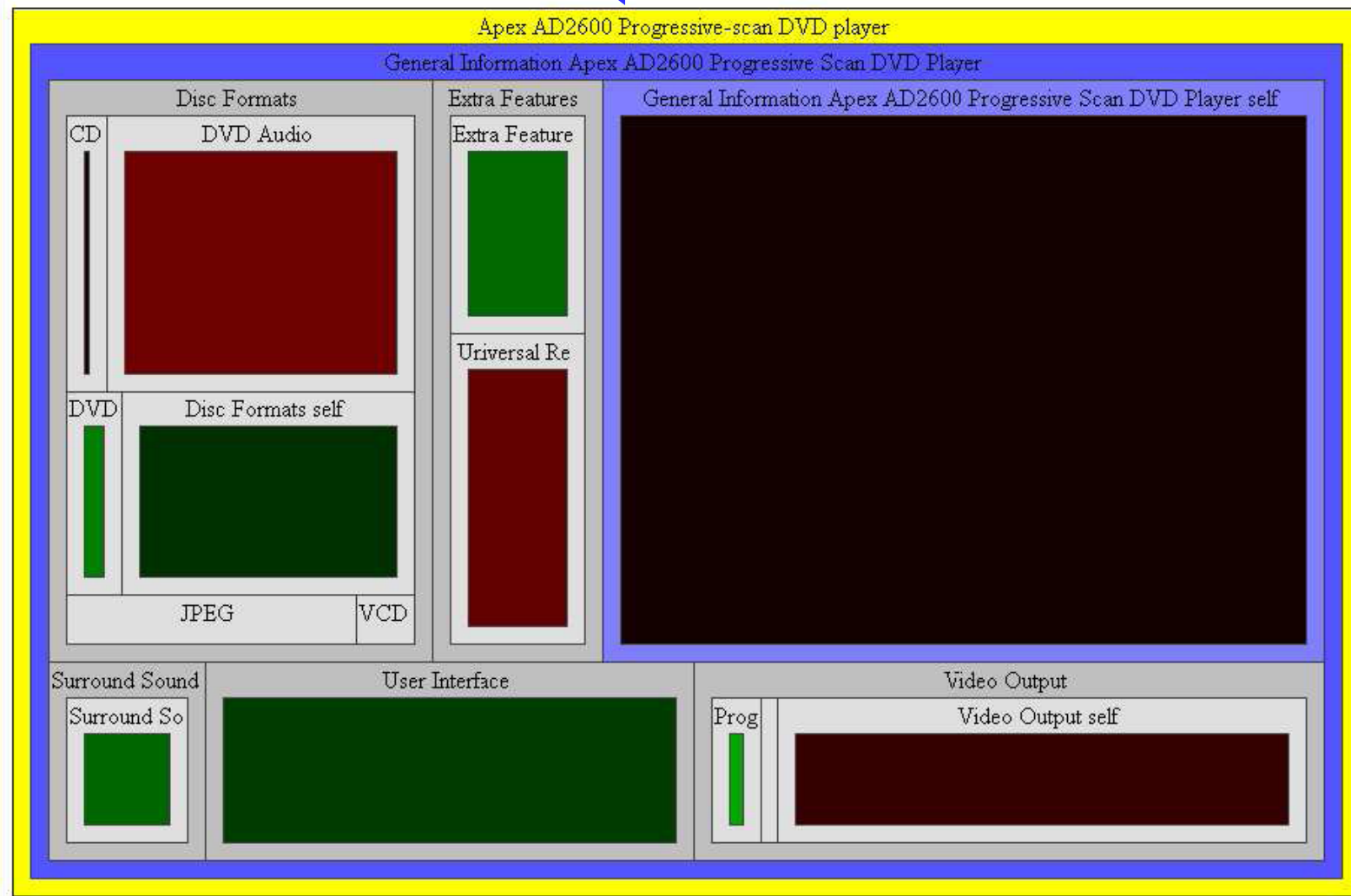


Textual Summary

Graphical Summary

Summary of customer reviews for: Apex AD2600 Progressive-scan DVD player

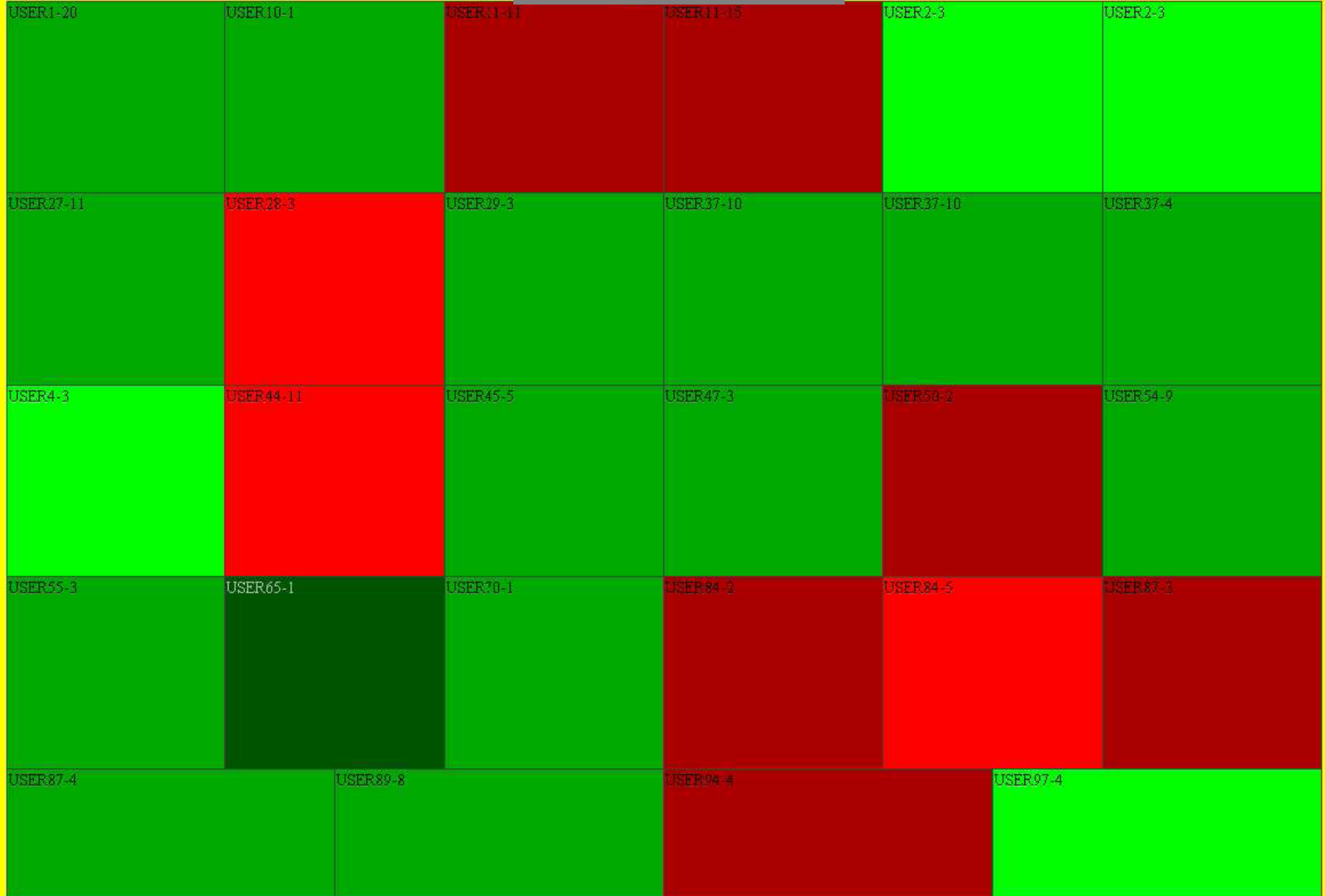
Most customers disliked the Apex AD2600 ¹. Although many customers found the user interface ² to be good, many users thought the available video outputs ³ was poor. However, many users liked the range of compatible disc formats ⁴, even though many customers found the compability with DVD audio ⁵ discs to be very poor.



For the price , it 's a very nice dvd player . The front door is miss aligned on my unit and you have to manually life it up just so slightly for the door to close , a very annoying thing after ahwile . **It does play a wide range of formats as advertised which is very nice .** And so far have not had any problems with dvds not being able to play . Recommended to anyone looking to purchase a low priced dvd player and not expecting any bells or whistles from a brand name one like sony .

← **Original Review(s)**

User Interface



Multimedia Interactive Summary: Formative Evaluation

- **Procedure** (similar to study-1 and study-2)
- Interested in testing effectiveness of text graphics combination (redundancy / support)
- Very positive feedback (Details in IUI-06 paper)
- **Recent Extension** to comparison of two entities (see paper in IUI-09)

Talk Summary

*Corpus of
Evaluative
Documents*

[KCAP '05]
Extract evaluative info:
feature hierarchy annotated
with evaluations

SEA (NLG abstractor)
MEAD* (extractor)

Generate
NL Summary

[EACL '06]

[INLG '08]

Allow access to
original sources:
•Text footnotes
•Treemap zooming

Present
As Treemaps

Present
NL Summary

Multimedia Summary

Interactive

[IUI '06]

[IUI '09]

Questions 2015

- UDFs / CFs / Gold Standard
- Unplaced CFs
- Clarification Placement distance
- CF extraction and polarity ←
- Constructing large UDF ←
- Different Languages
- Trade-off Placement and Redundancy
- Future
 - MSR ←
 - Interactive Multimedia Summarization ←
 - Extract Hierarchy from the reviews (automatically create UDFs)..... Speech input .. Sarcasm

Placement Distance

- The accuracy of a CF term in the research is assessed by considering the hierarchical path distance between where it is placed by the mapping algorithm and where it is placed by the gold standard (control mapping). Does the research assume that path lengths all encode the same semantic distance? (e.g. that pixels (parent) -> resolution (child) has a semantic subset distance equal to image (parent) -> image type (child))

- G. Carenini , J. Cheung , A. Pauls. [Multi-Document Summarization of Evaluative text](#), **Computational Intelligence, 2012**
- Carenini G. and Rizoli L., A Multimedia Interface for Facilitating Comparisons of Opinions. . In: *Proceedings of the 13th International Conference on Intelligent User Interfaces, (IUI 2009)*, Sanibel Island, Sydney, Florida, USA, 2009 [[pdf](#)]
- Carenini G, Cheung J., Extractive vs. NLG-based Abstractive Summarization of Evaluative Text: The Effect of Corpus Controversiality. *International Conference on Natural Language Generation. (INLG 2008)*, Salt Fork, Ohio, USA, June 12-14, 2008 [[pdf](#)]

Carenini G., Ng R., Pauls A. MultiDocument Summarization of Evaluative Text, *Proceedings of the 11th European Chapter of the Association for Computational Linguistics (EACL 2006)*, Trento, Italy, April, 2006. [[pdf](#)]

Carenini G., Ng R., Pauls A. Interactive Multimedia Summaries of Evaluative Text, *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2006)*, Sydney, Australia, Gen29-Feb1, 2006. [[pdf](#)]

Results for DVD

Table 2: Placement distance and redundancy reduction scores for DVD player with term metric *avg*

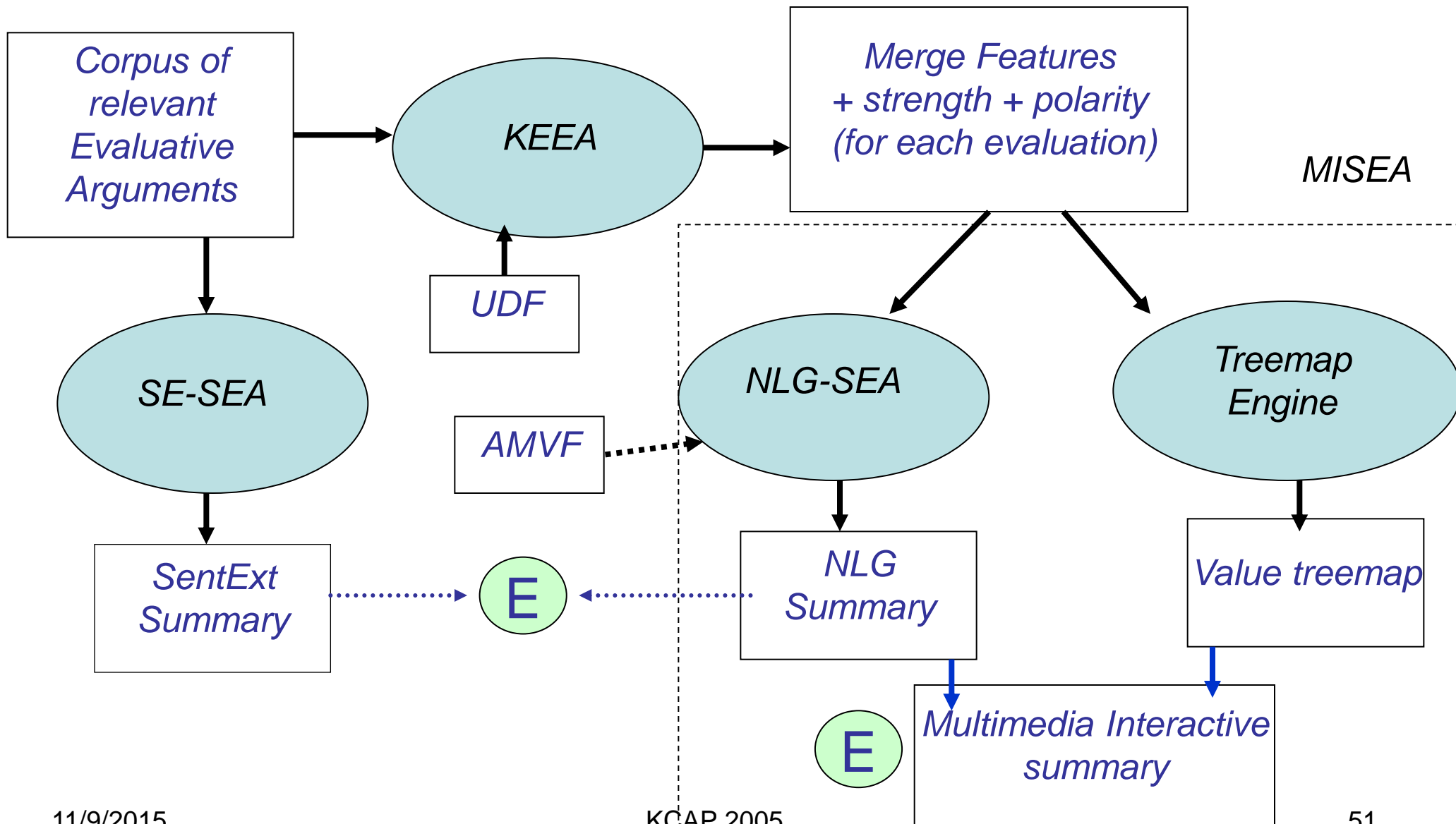
	1st Run		No Repetition	
	p_dist	redun	p_dist	redun
str_match	.31	.19	.27	.21
syn_score	.30	.23	.28	.25
θ	sim_score (res)			
-0.2	.39	.30	.39	.32
-0.4	.49	.44	.49	.46
-0.6	.59	.53	.58	.54
θ	sim_score (lin)			
-0.2	.42	.36	.40	.38
-0.4	.49	.43	.47	.45
-0.6	.57	.50	.55	.52

Results DigCam for AVG

Table 1: Placement distance and redundancy reduction scores for DigCam with term metric *avg*

	First Run		No Repeat	After Revision
	p_dist	redun		
str_match	.43	.19		
syn_score	.45	.21		
θ				
-0.2	.42	.16		
-0.4	.42	.23		
-0.6	.46	.31		
θ				
-0.2	.44	.23		
-0.4	.43	.31		
-0.6	.43	.36		

Support Analysis of Evaluative Arguments (about single entity)



Results: Summary

- In both products Wordnet distance with $\theta = -.2$

From Mapping To Evaluation

- Given unsupervised CF extraction and unsupervised UDF \leftrightarrow CF mapping, need to evaluate UDF features
- Assume we can calculate strength and polarity of customer evaluations for each CF using existing methods (Hu & Liu 2004; Wilson *et al.* 2004),
 - then we can generate an evaluation for each UDF based on its CF's

Back to High-level process

- Information Extraction
- **Summary generation**

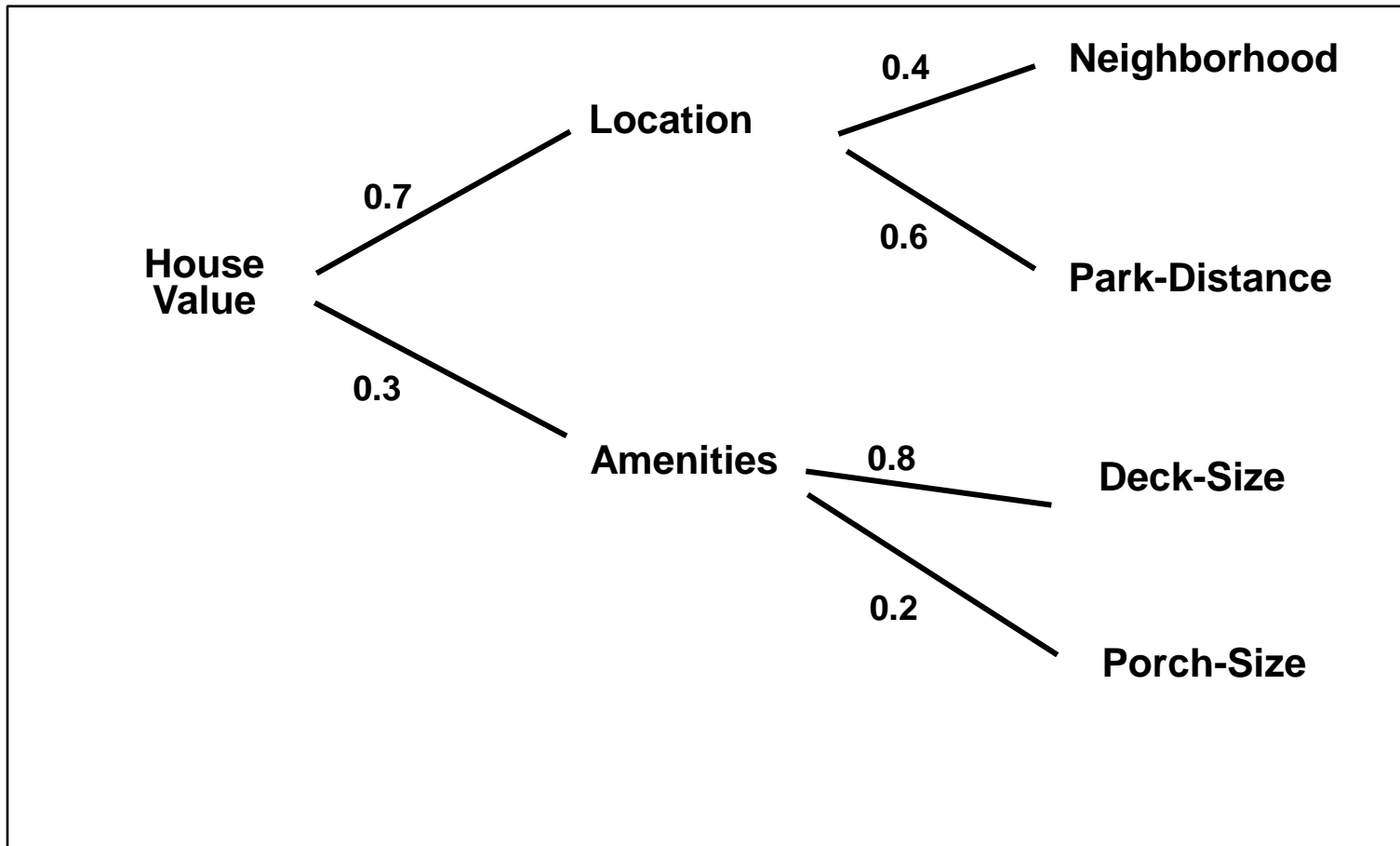
Plan for Summary Generation

- **Adapt GEA (Generator of Evaluative Arguments) (Carenini & Moore 2001) for**
 - Content selection and organization
 - Microplanning (partially)
 - Realization
- Adapt existing MEAD (Radev et. al. 2001) software as baseline “domain/task independent” summarizer
- Evaluation: Compare system against baseline with human judges

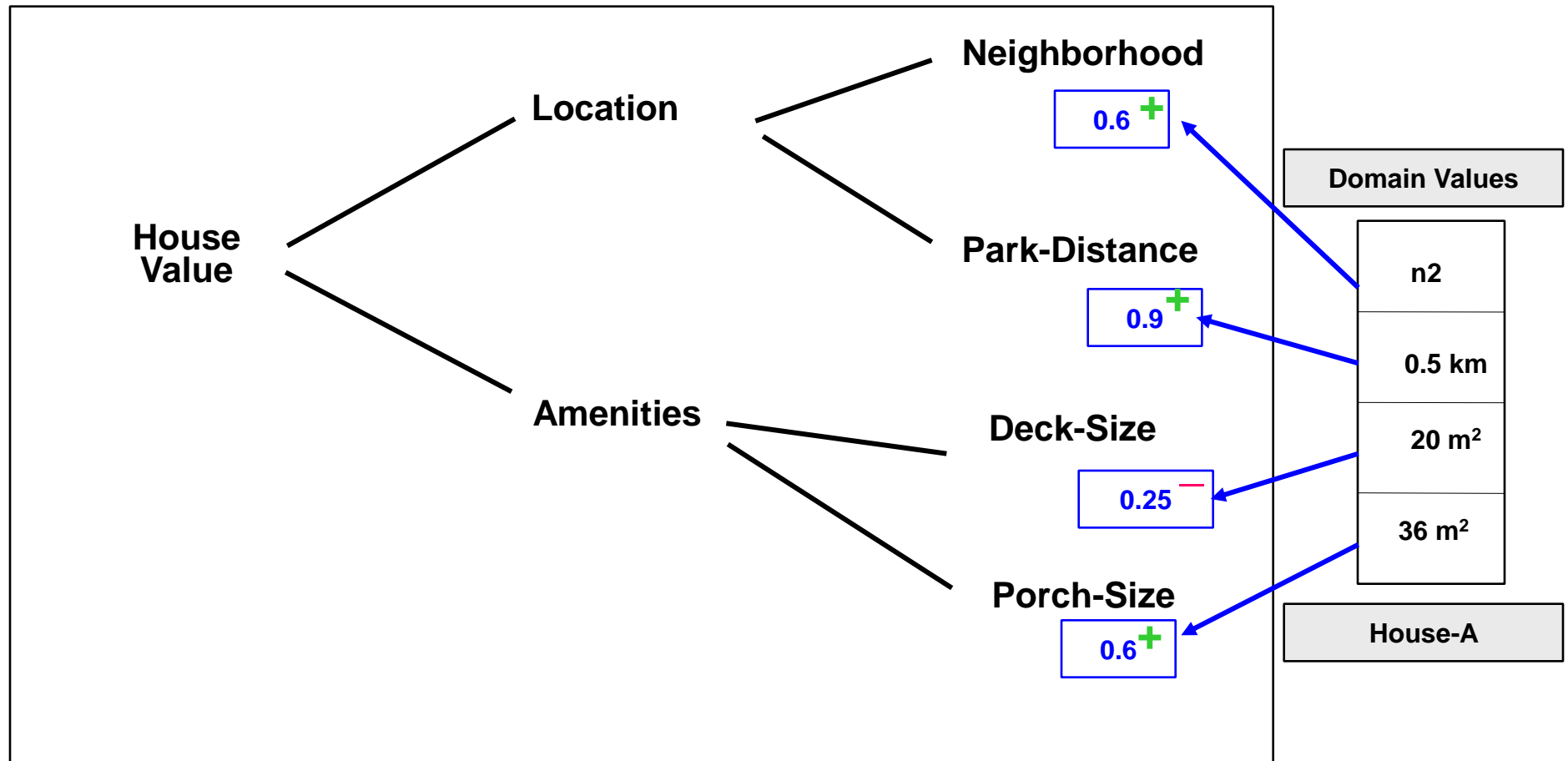
Generator of Evaluative Arguments (GEA)

- Generates evaluations of entities based on:
 - properties of entity
 - user preferences about that entity
- *Entity is represented as a set of attributes and values (e.g. (Zoom range . 12x))*
- *User Preferences are modelled using an AMVF (**A**dditive **M**ultiattribute **V**alue **F**unction)*
 - *This is a hierarchical set of preferences about entity, with attributes as leafs*

GEA example: AMVF



GEA example: Attributes



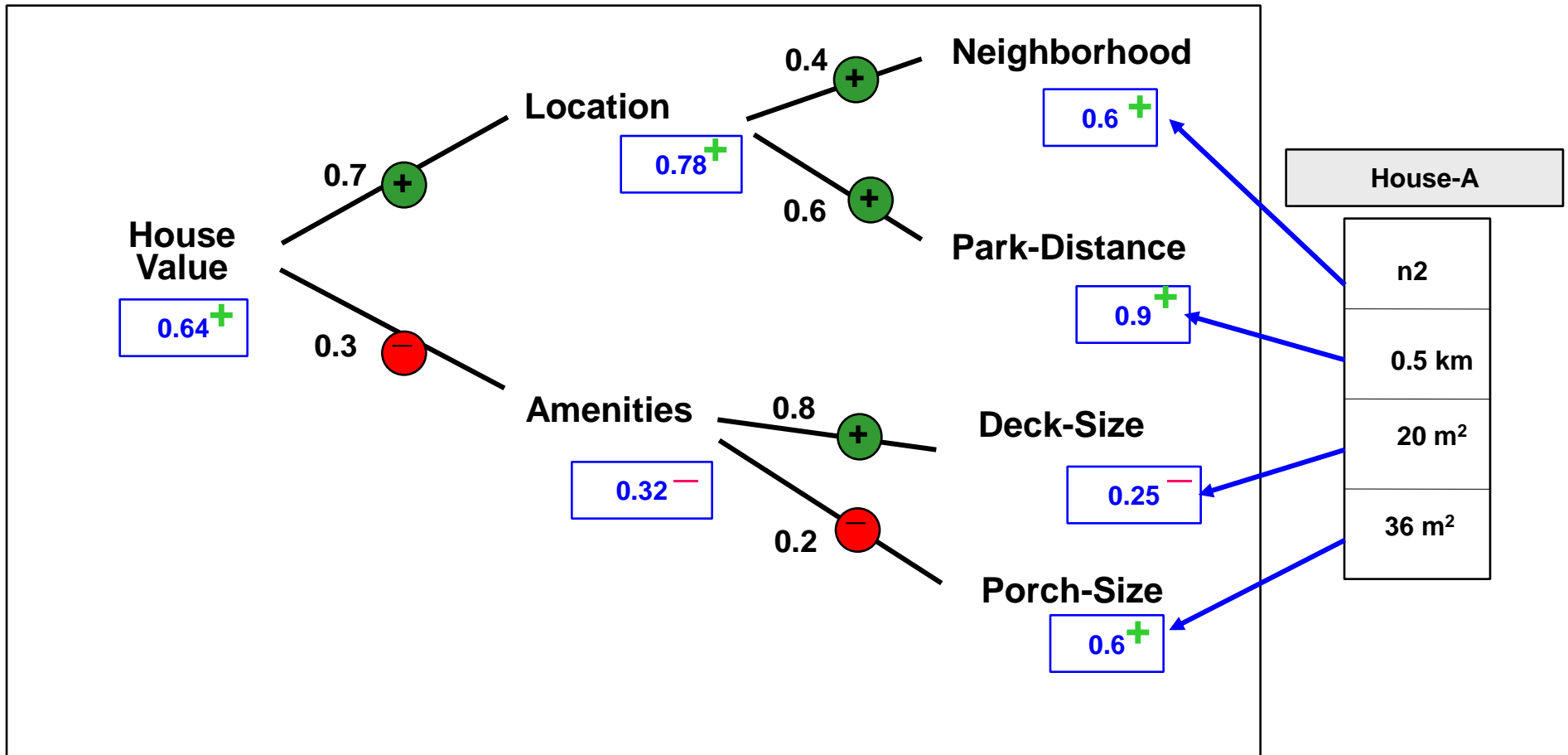
+	Likes it
-	Does not like it



Component Value

Function

GEA example: Opposing/Supporting Evidence



+	Likes it
-	Does not like it

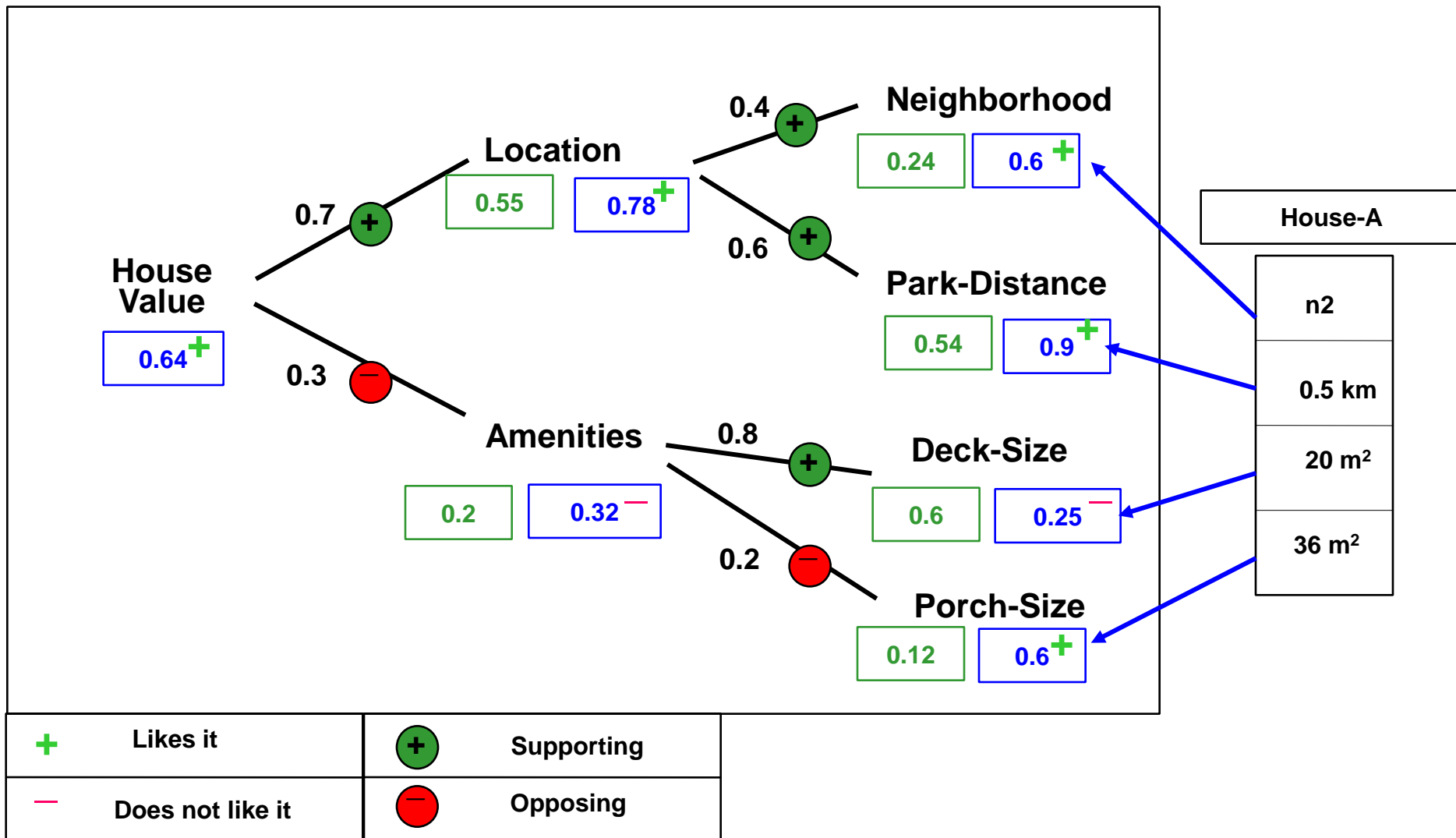
<i>o</i>	<i>Parent(o)</i>	<i>relation</i>
+	+	supporting
-	-	supporting
+	-	opposing
-	+	opposing

+	Supporting
-	Opposing

Measure of Importance [Klein 94]

For each attribute a :

$$\text{Importance}(a) = w_a \max[v_a, 1 - v_a]$$



Argumentative Strategy

Based on guidelines from argumentation theory

[Miller 96, Mayberry 96]

Selection: include only “important” evidence
(i.e., above threshold on measure of importance)

Organization:

- (1) **Main Claim** (e.g., “This house is interesting”)
- (2) **Opposing evidence**
- (3) **Most important supporting evidence**
- (4) **Further supporting evidence -- ordered by importance with *strongest last***

Strategy applied recursively on supporting evidence

Adapting GEA

GEA

- AMVF hierarchy ->
- AMVF weights ->
- Component Value Function ->

Customer Reviews

- UDF hierarchy
- Relative frequency of UDFs in corpus
- Aggregation of polarity/strength of UDF features

Adapting GEA (cont'd)

- Differences
 - **Customers may evaluate non-leaf elements (e.g. “Location”) directly**
 - **in GEA domain, entities had only one evaluation for each attribute**
 - **For customer reviews, must give some indication of distribution of customer opinions on each attribute**

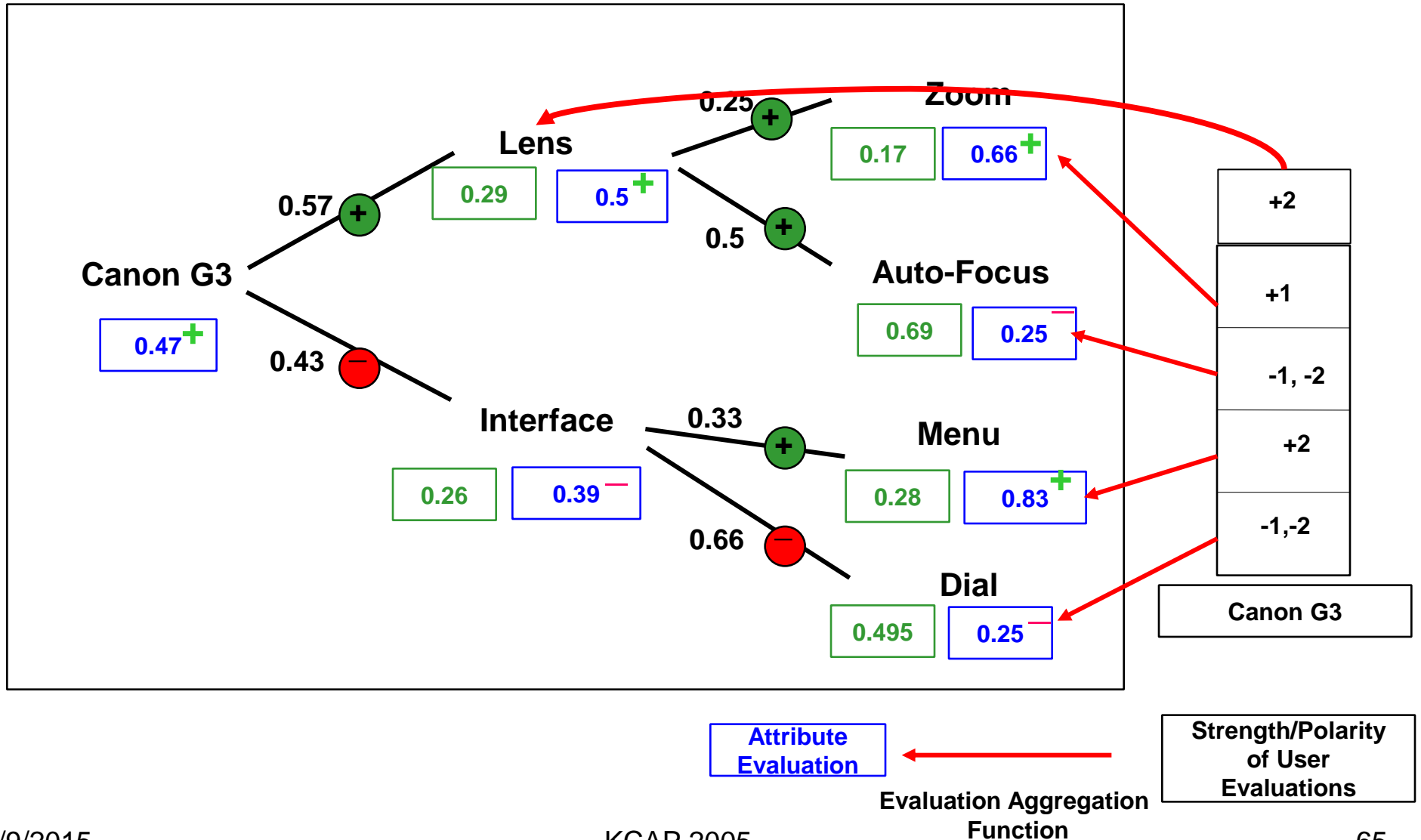
Example: Some (fake) Reviews

"I really liked the Canon G3 **[+2]**. The 12x zoom is really useful **[+1]**! The only thing I didn't like was its poor **[-1]** focussing in low light."

"The Canon G3 is a great deal. The lens features were the best I've seen for a camera of its price **[+2]**. The menu system is very intuitive **[+1]**, but I wish the camera could take RAW images **[-1]**."

"I really didn't like this camera **[-2]**. It focussed very poorly **[-2]** indoors (when I use it most) and I found myself wishing there were more modes on the dial **[-1]** rather than in the menu system. I returned mine already."

Adapted GEA



Output of GEA

- What GEA gives us:
 - High-level text plan (i.e. content selection and ordering)
 - Cue phrases for argumentation strategy (“In fact”, “Although”, etc.)
- What GEA does not give us:
 - Appropriate micro-planning (lexicalization).
 - Need to give indication of distribution of customer opinions

Hypothetical GEA Output

The Canon G3 is a good camera. However, the interface feature is poor. Although the menu system is good, the dial system is terrible.

Target Summary

Most users thought Canon G3 was a good camera.
However, several users did not like
interface. Although most users liked the menu
system, many thought the dial was terrible.

Evaluation

- Current idea: *task-based (extrinsic) evaluation*
 - Give human test subject summary
 - Then, allow user some fixed time (e.g. 5 minutes) to scan a corpus of reviews (20-30?)
 - User should then answer e.g.
 - if summary provides “all” (?) important information
 - if summary left out information
 - if missing information was important
 - if summary is representative of corpus
 - Also evaluate fluency with known methods

Future Directions

- Current method of adapting GEA is just a first pass.
 - Could change e.g. Measure of Importance.
- We may leverage GEA's ability to create *user-tailored evaluative arguments for generating user-tailored summaries (long term)*

IE Key Sub-tasks

- A.** What **features** of the objects are evaluated in the reviews?
- B.** For each feature:
 - i. what is the **polarity** of the evaluation? (good vs. bad)
 - ii. what is the **strength** of the evaluation? (rather good vs. extremely good)

(User-Specific) Summarization of Multiple Customer Reviews

The Goal:

An automatic solution to the problem of **summarizing a potentially large set of documents that contain evaluative language about a given entity** (e.g., a product, a location, a job candidate, etc.)

User Specific: the summary is tailored to user's conceptualization of the entity (now)
model of the user's preferences (long term)

Example: Some (fake) Reviews

"I really liked the Canon G3. The 12x zoom is really useful! The only thing I didn't like was its poor focussing in low light."

"The Canon G3 is a great deal. The lens features were the best I've seen for a camera of its price. The menu system is very intuitive, but I wish the camera could take RAW images."

"I really didn't like this camera. It focussed very poorly indoors (when I use it most) and I found myself wishing there were more modes on the dial rather than in the menu system. I returned mine already."

Example: Target Summary

Most users liked the Canon G3. Many found the zoom feature to be good. Although many users did not like the auto focus, a few users liked the menu system. Only 1 user did not like the camera.

Example Target Summary

- Features
 - *Selection of content (flash range not mentioned)*
 - *Discourse cues (cue phrases, order of evidence)*
 - *Contrasting and supporting evidence for summary of camera*
 - *Lexicalization of numerical tallies (2/3 => “most”)*

High-level process

- **Information Extraction**
- Summary generation

Example of Learned Features for a Digital Camera

- noise
- function
- button
- camera
- four megapixel
- remote control
- software
- manual
- remote
- lever
- price
- Canon G3
- strap
- low light focus
- tiff format
- use

Ideal Extraction: sample form corpus

[Hu&Liu 2004]

..... the canon **computer software** [+2]
used to download , sort , . . . is
very nice and very easy to use. the
only two minor issues i have with
the camera are the **lens cap** [-1] (
it is not very snug and can come
off too easily)

the **menus** [+1] are easy to navigate
and the **buttons** [+1] are easy to
use. it is a fantastic **camera** [+3]
and well worth the price .

Where we are now...

- Increase in accuracy scores from measure to measure, and from MAX to AVG, but it's small.
- We need to understand better how the similarity measures are working to better take advantage of them.