SPAWAR

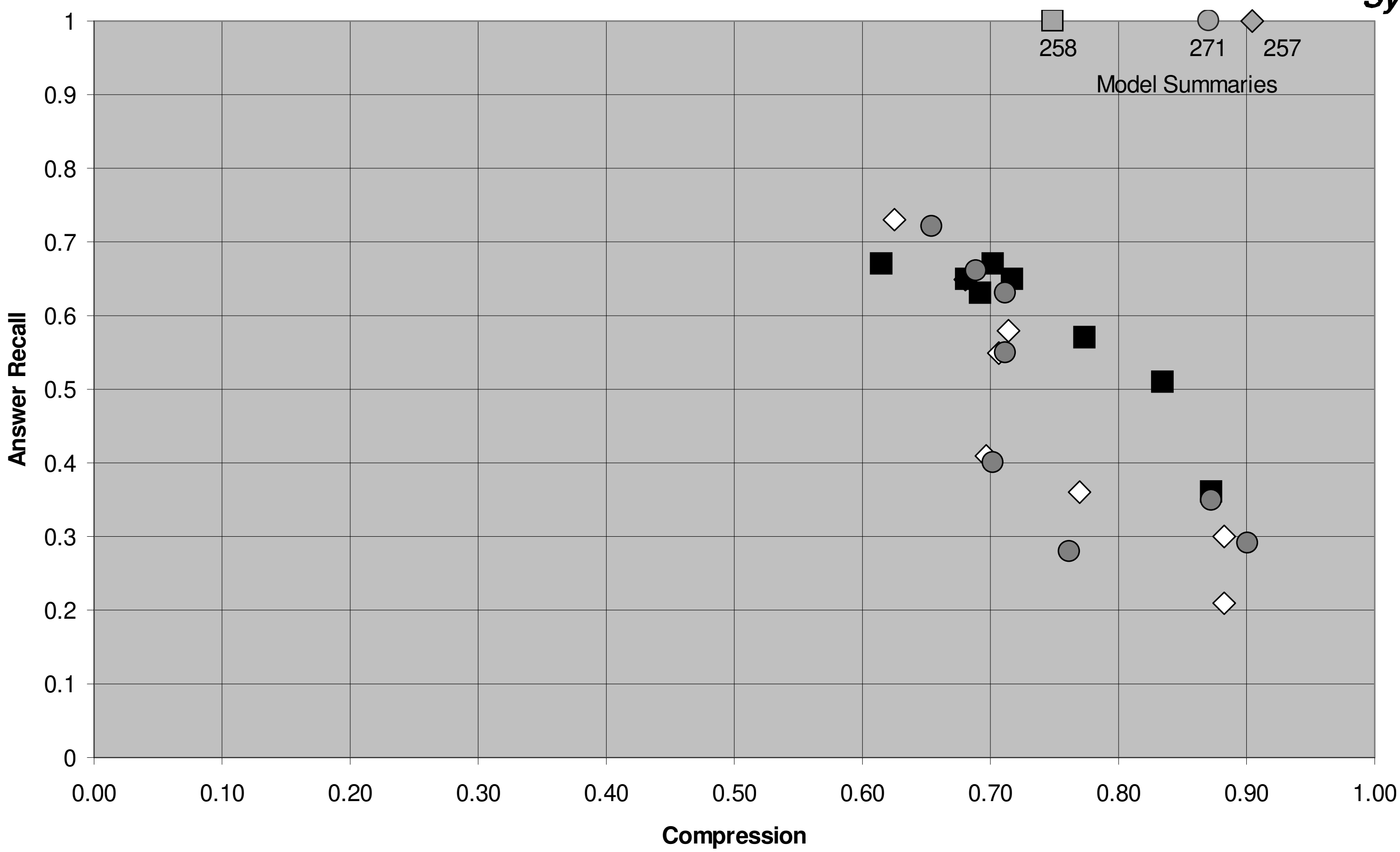
Systems Center
San Diego

The Tipster Question-and-Answer (Q&A) Summarization Task: Test Results

Beth Sundheim
sundheim@spawar.navy.mil
619/553-4145

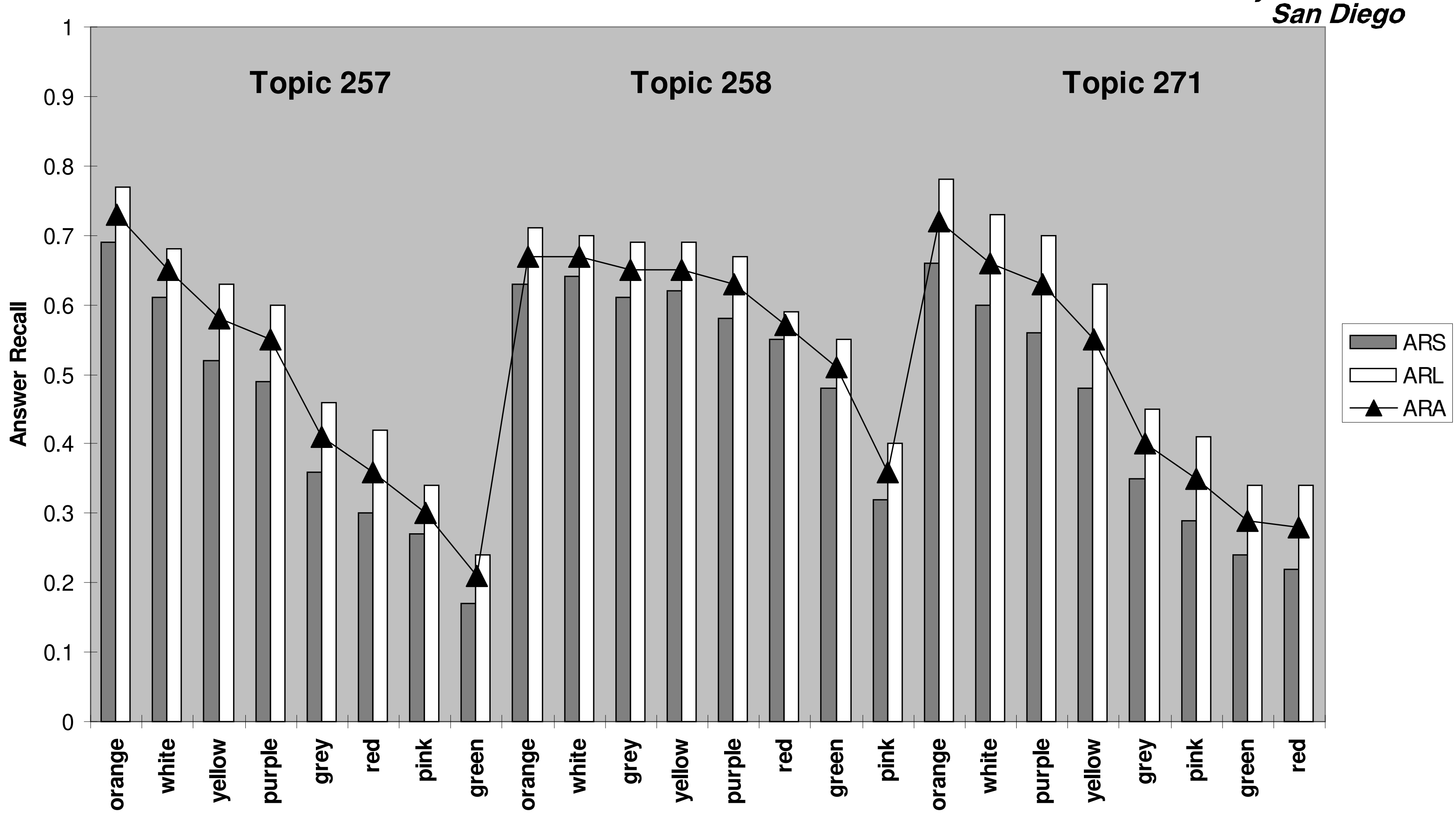
Answer Recall (ARA) vs Compression (1-Compression) for Each Topic and Site

Systems Center San Diego

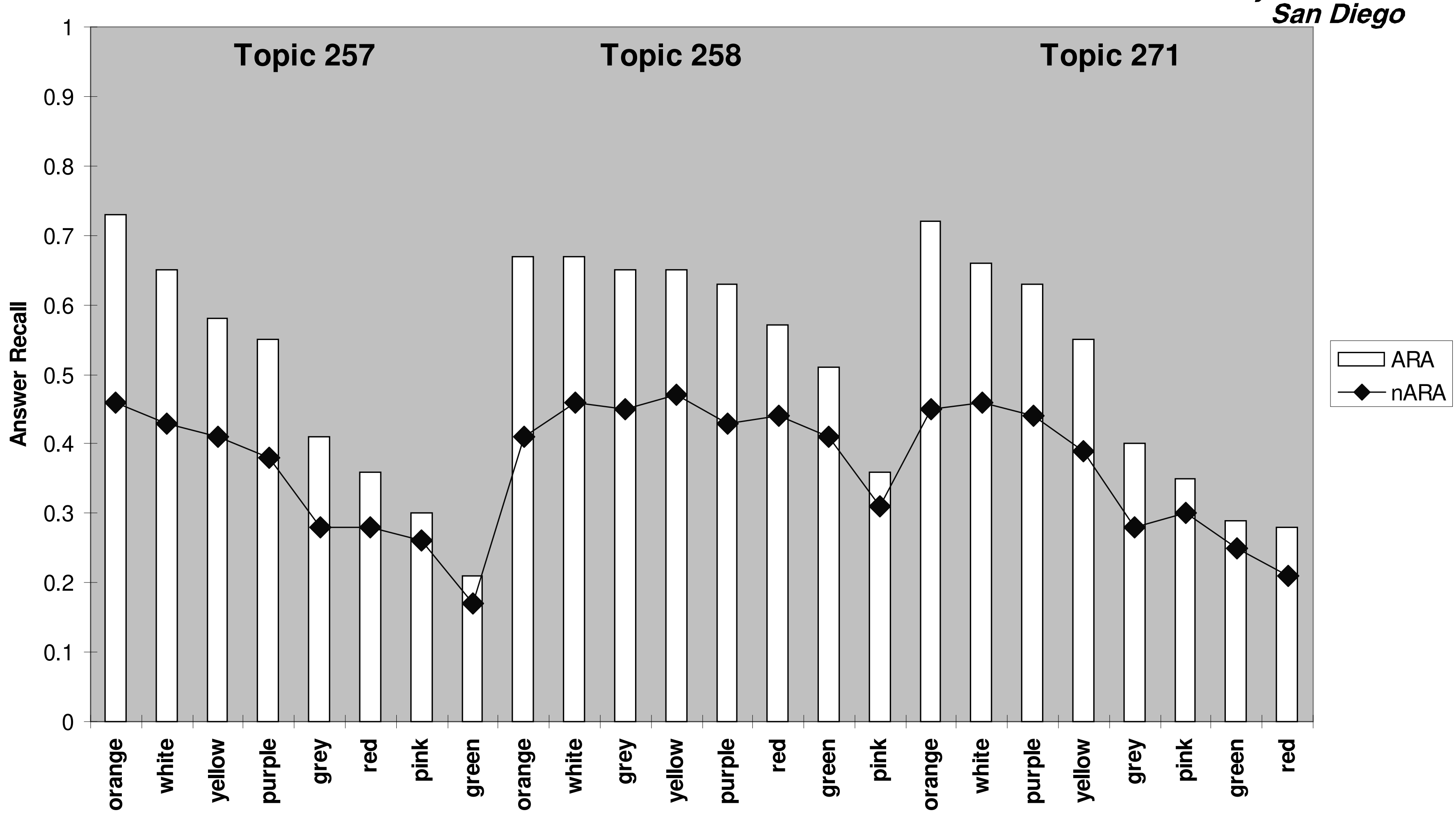


Unnormalized Answer Recall by Topic and Site (Sites Ordered by Decreasing ARA)

Systems Center San Diego

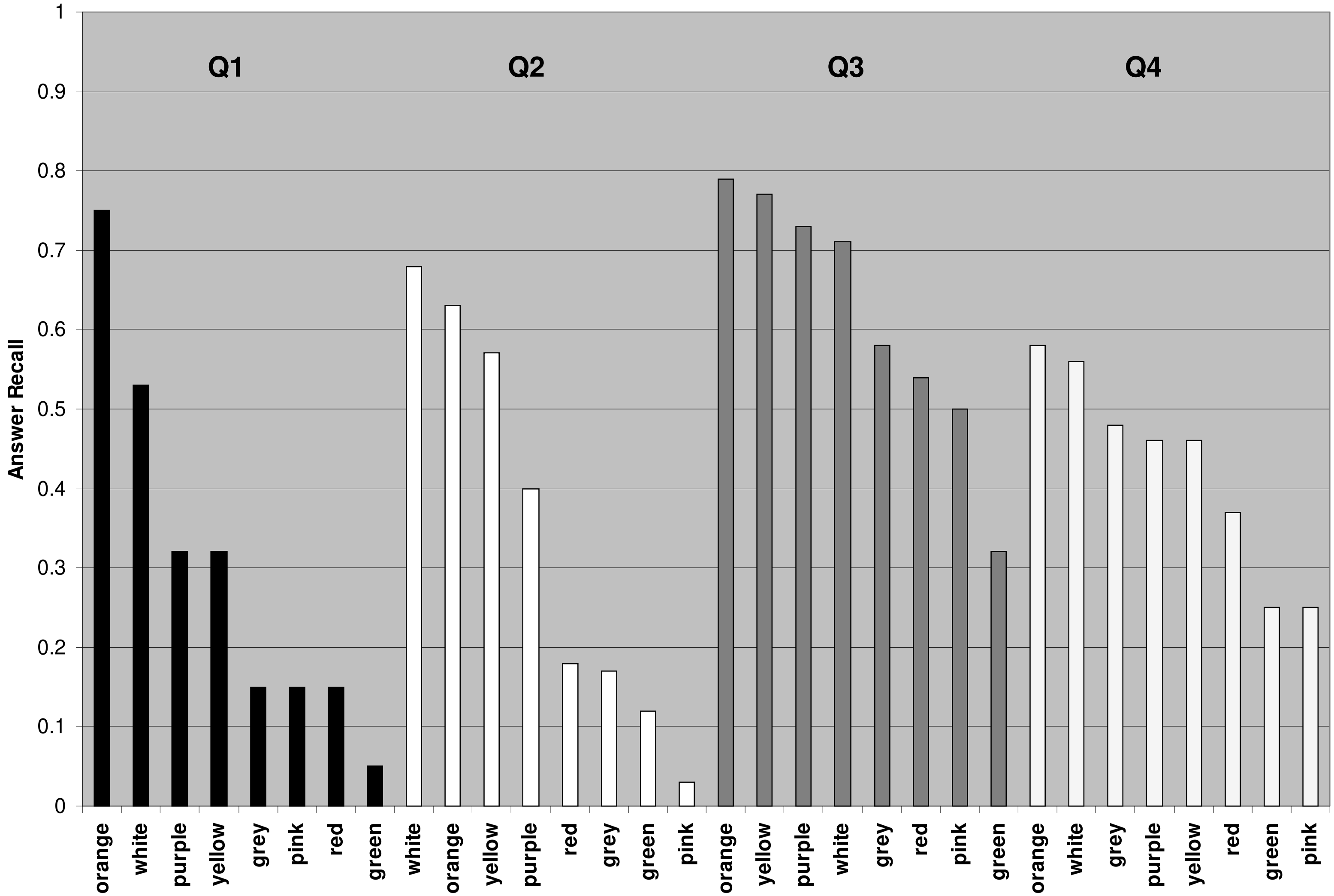


Unnormalized and Normalized Recall Average by Topic and Site (Sites Ordered by Decreasing ARA)



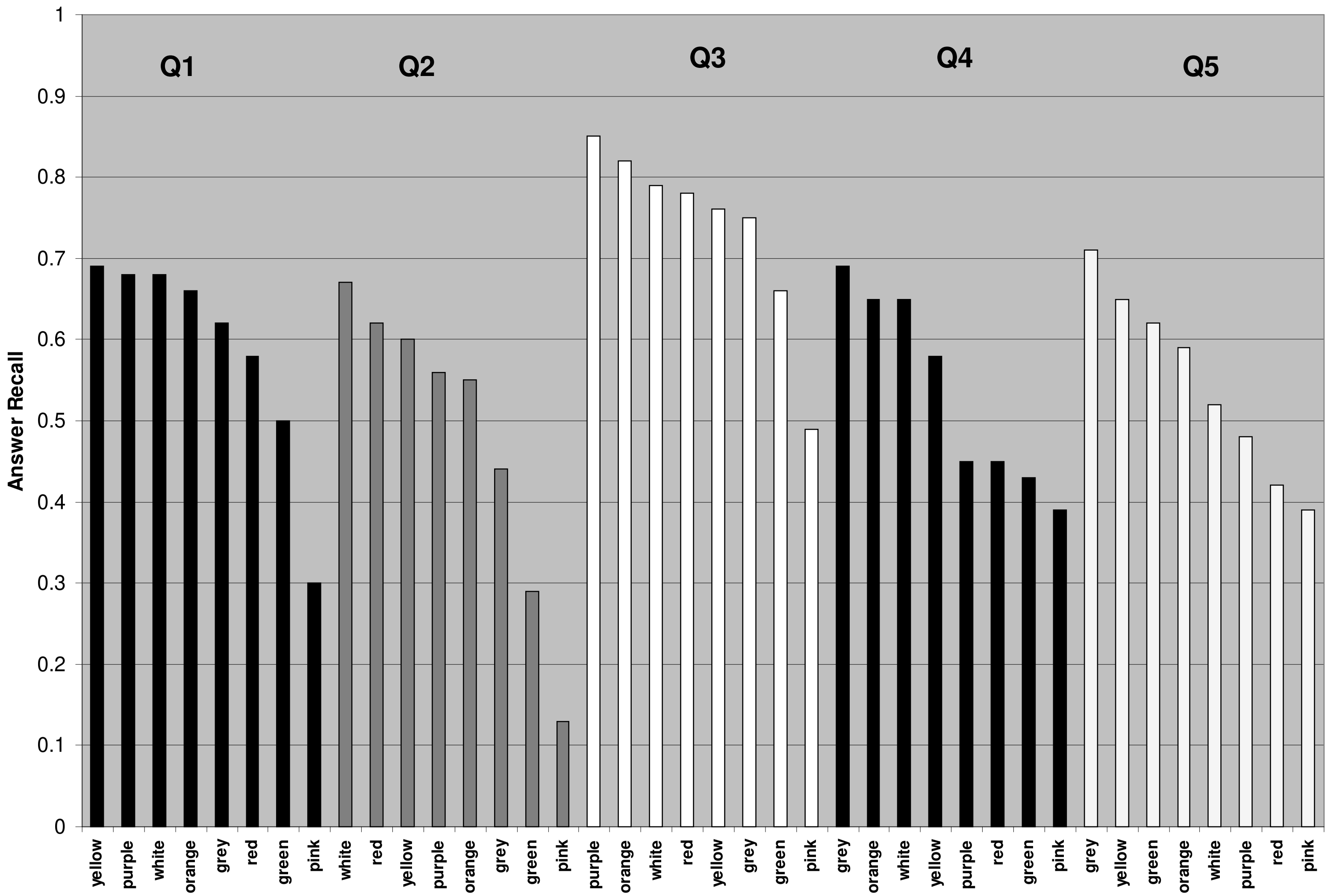
Answer Recall (ARA) by Question (Q1-Q4): Topic 257

SPAWAR Systems Center San Diego

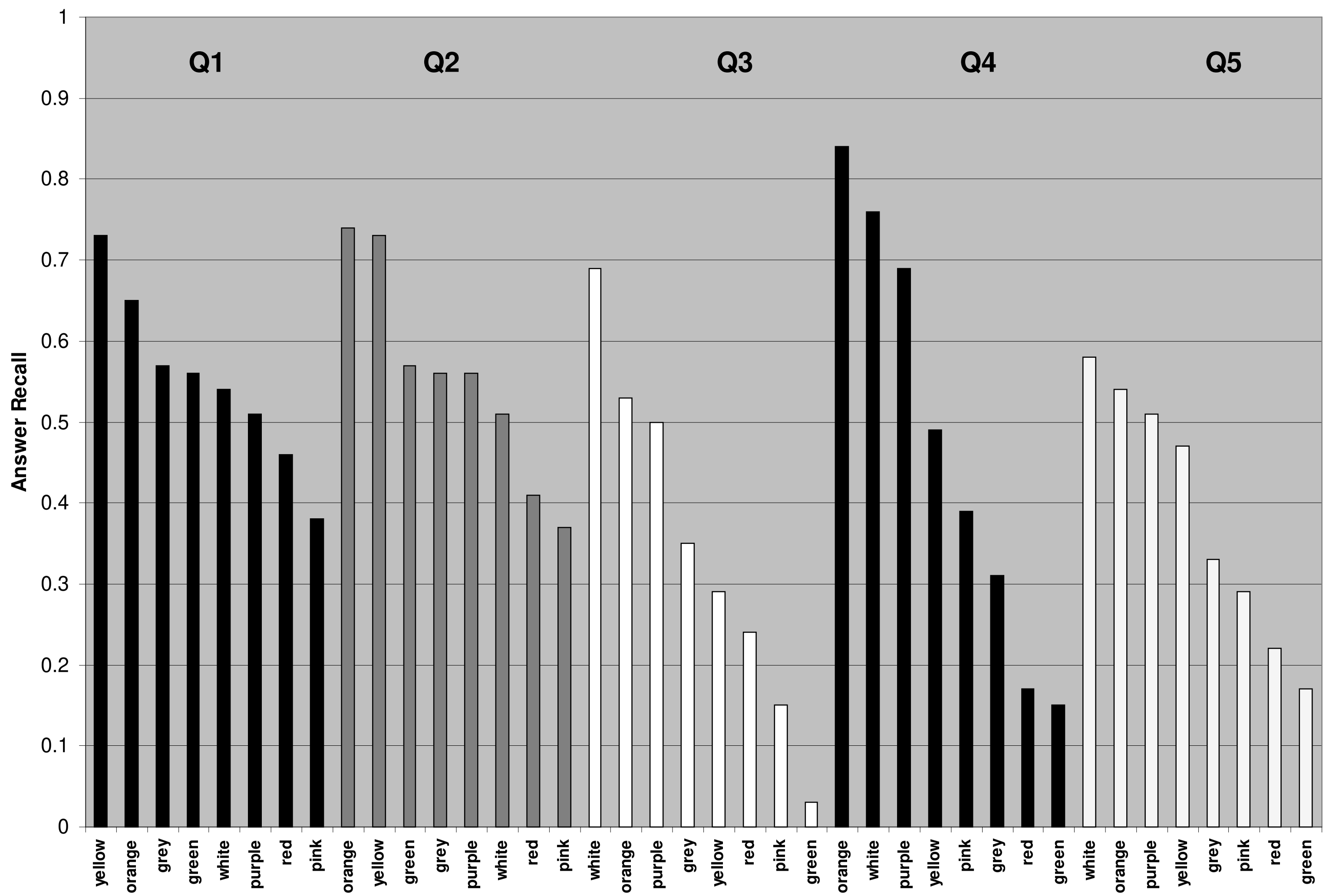


Answer Recall (ARA) by Question (Q1-Q5): Topic 258

Systems Center San Diego

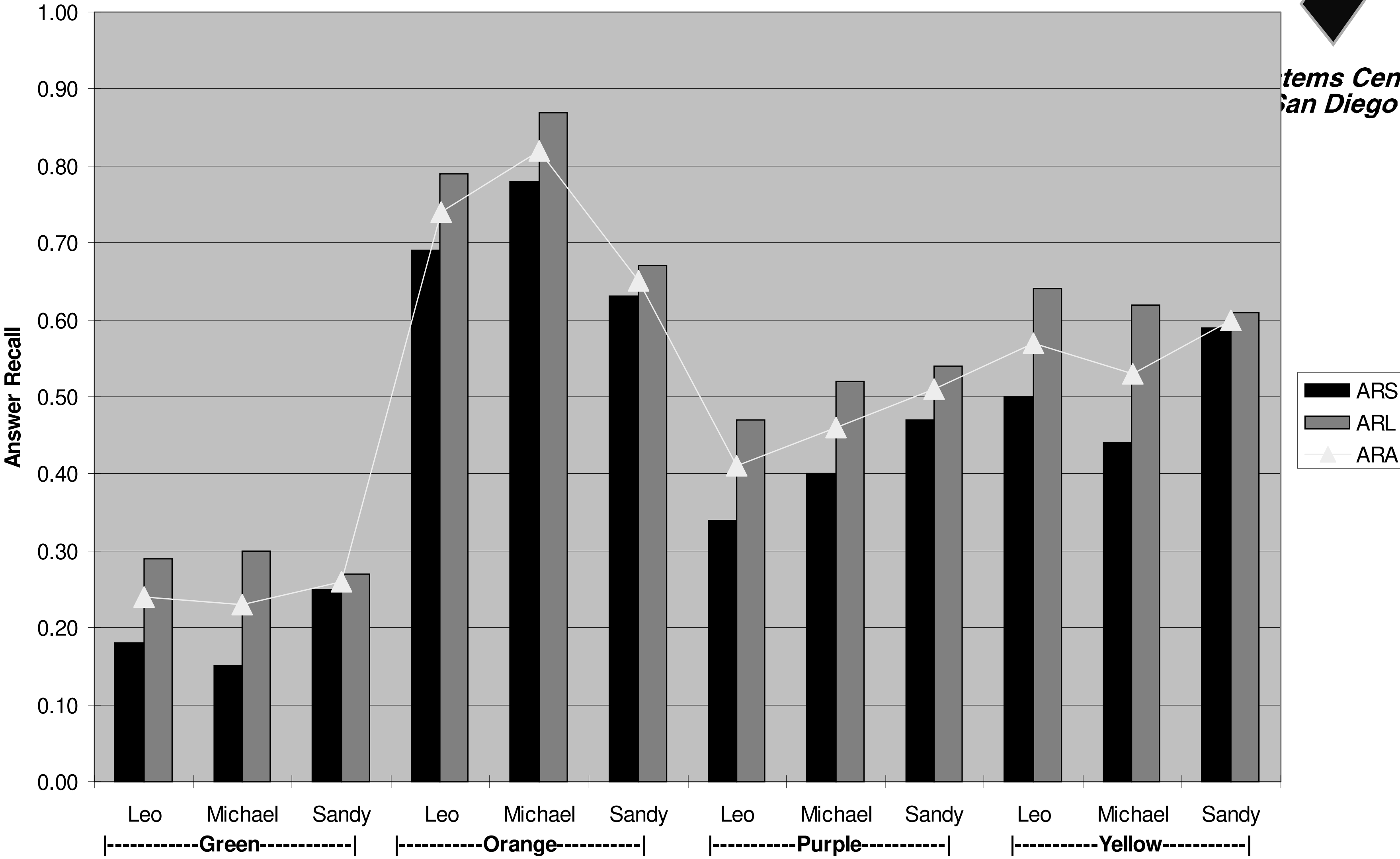


Answer Recall (ARA) by Question (Q1-Q5): Topic 271

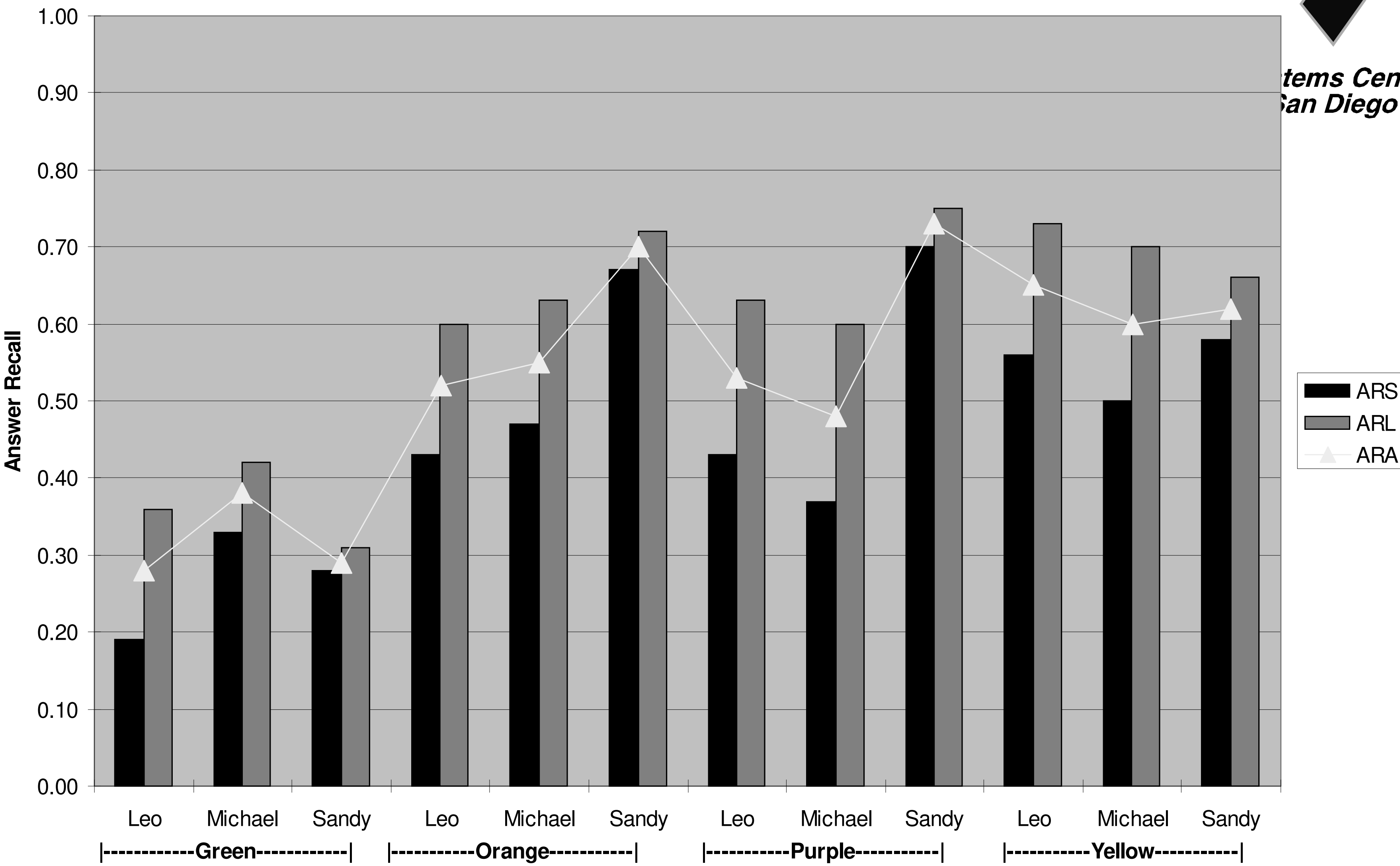


Systems Center San Diego

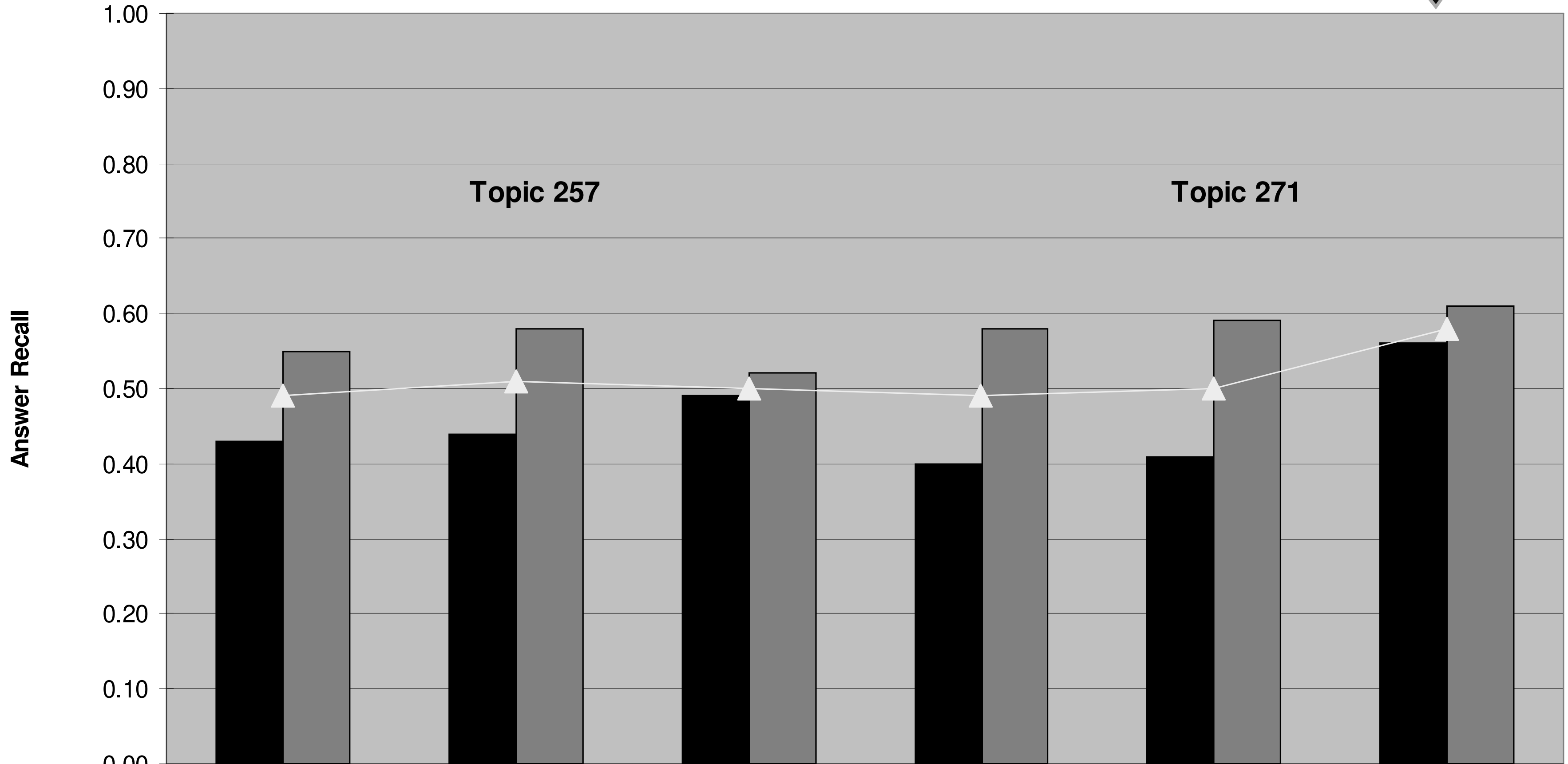
Interassessor Results by Site: Topic 257



Interassessor Results by Site: Topic 271



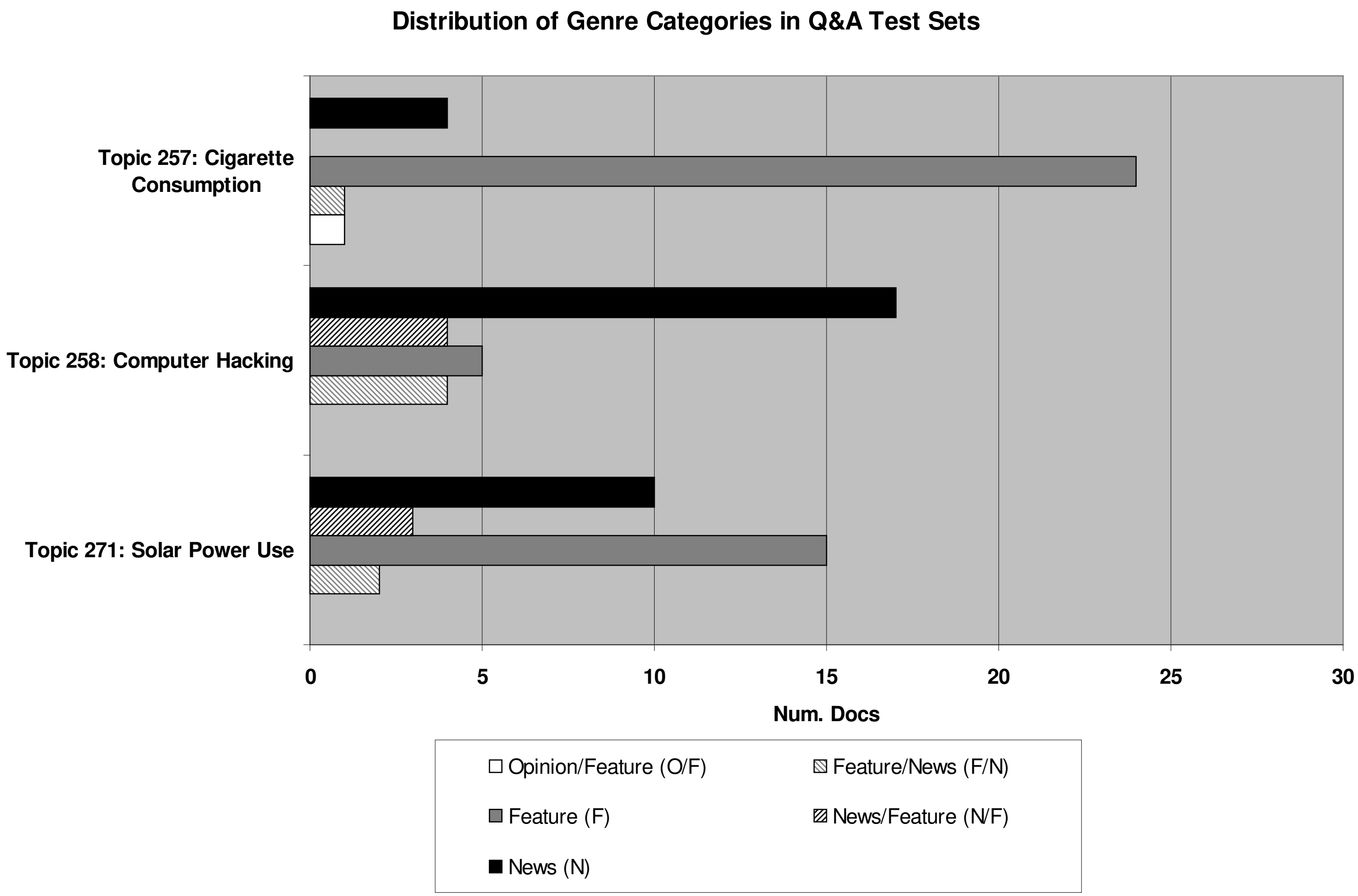
Interassessor Results: Summary by Topic



	Leo	Michael	Sandy	Leo	Michael	Sandy
ARS	**0.43**	**0.44**	**0.49**	**0.40**	**0.41**	**0.56**
ARL	**0.55**	**0.58**	**0.52**	**0.58**	**0.59**	**0.61**
ARA	**0.49**	**0.51**	**0.5**	**0.49**	**0.5**	**0.58**

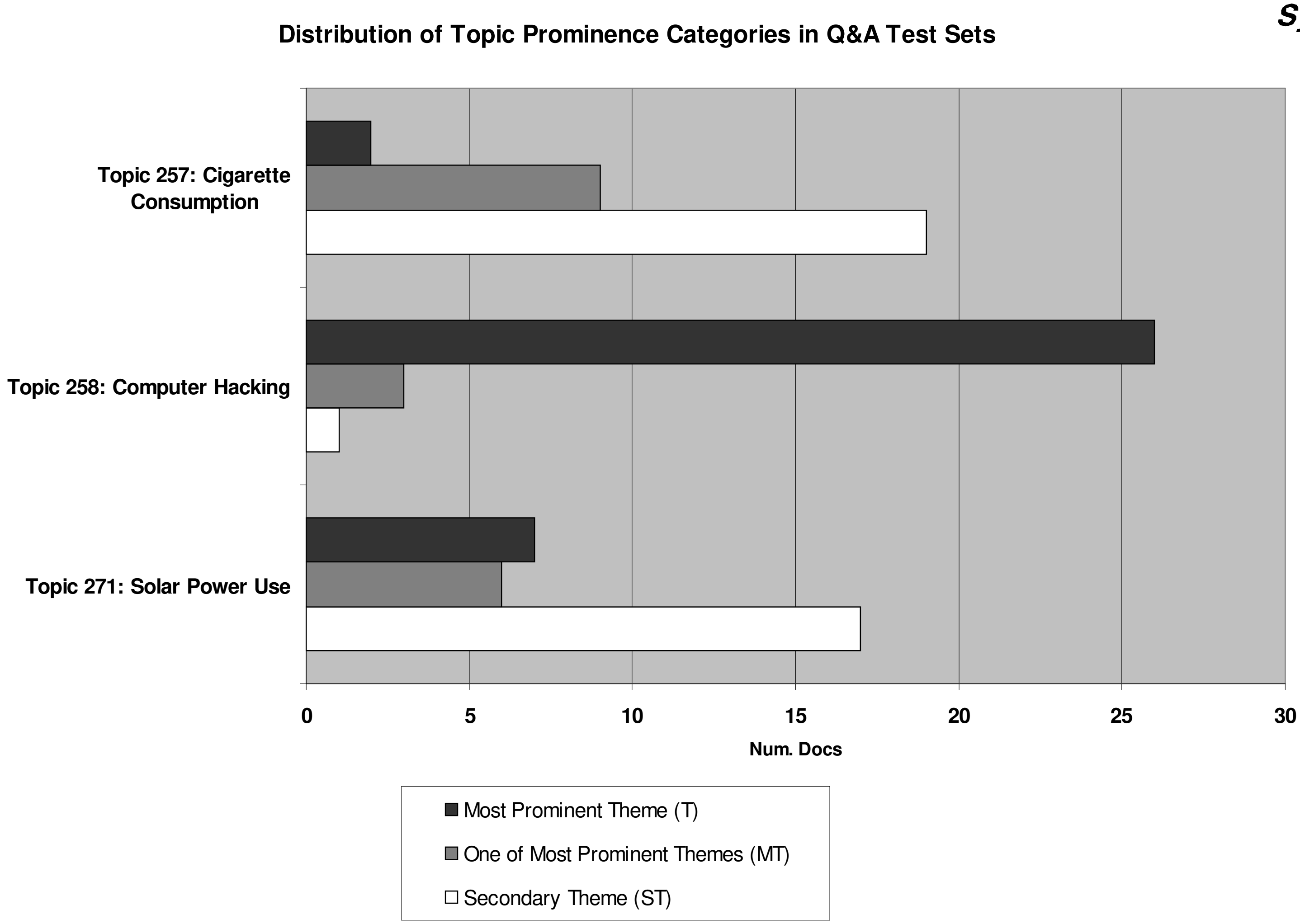
Cross-Topic Differences in Document Characteristics

Systems Center San Diego

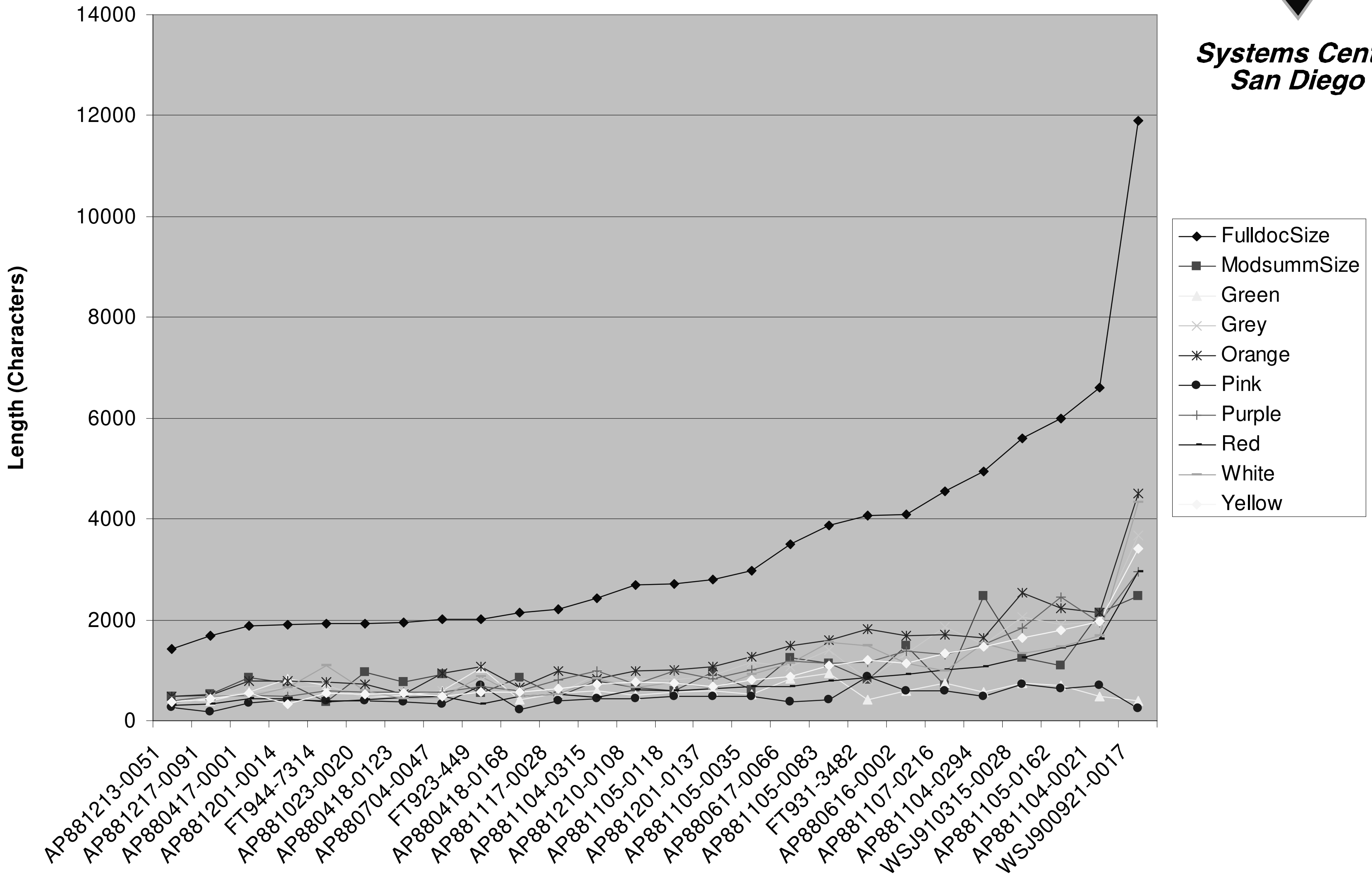


Cross-Topic Differences in Document Characteristics (Cont'd)

Systems Center San Diego

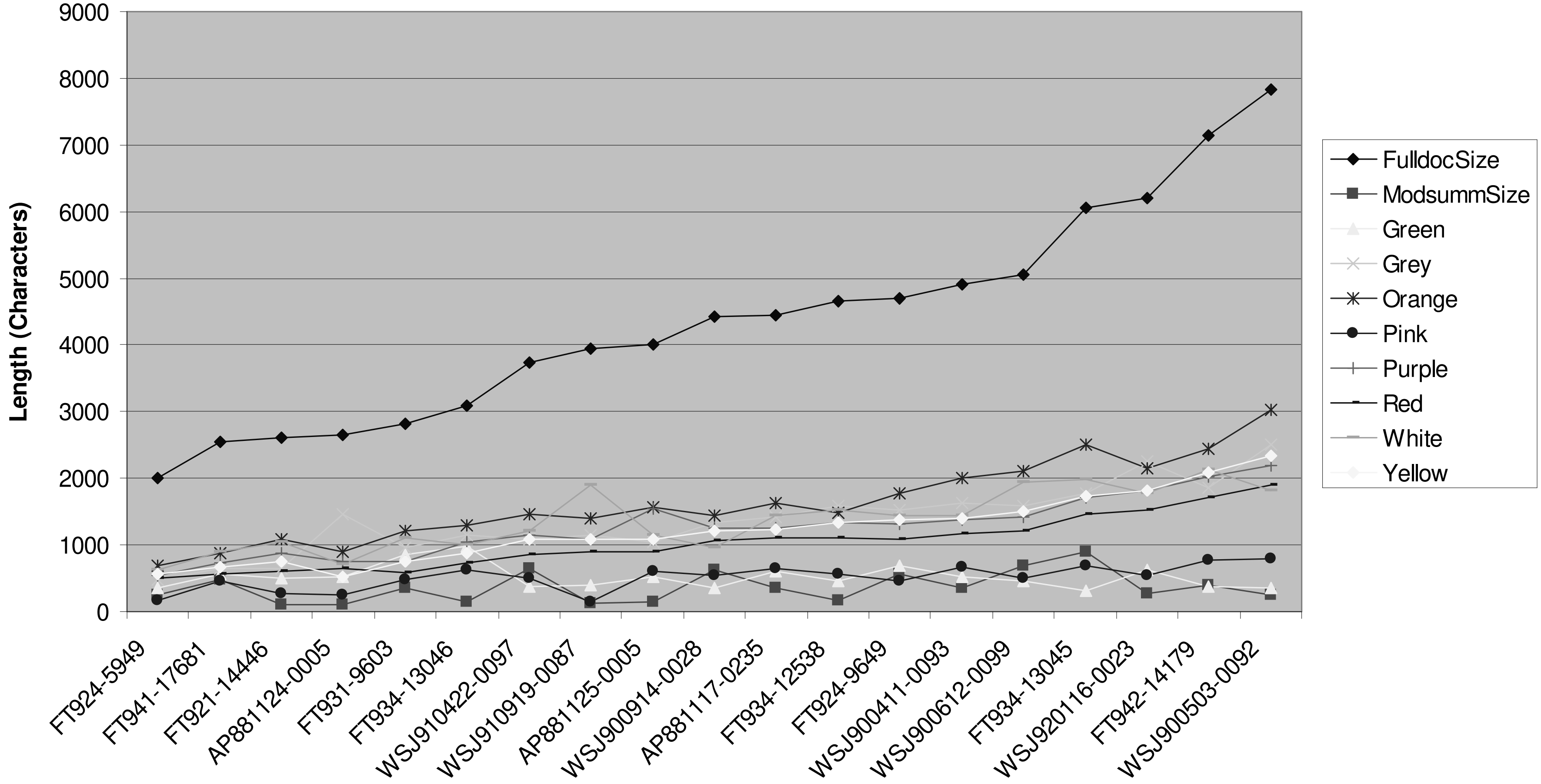


Length of Summaries in Relation to Full Docs for "Most Prominent Theme" (T) Category: Topic 258



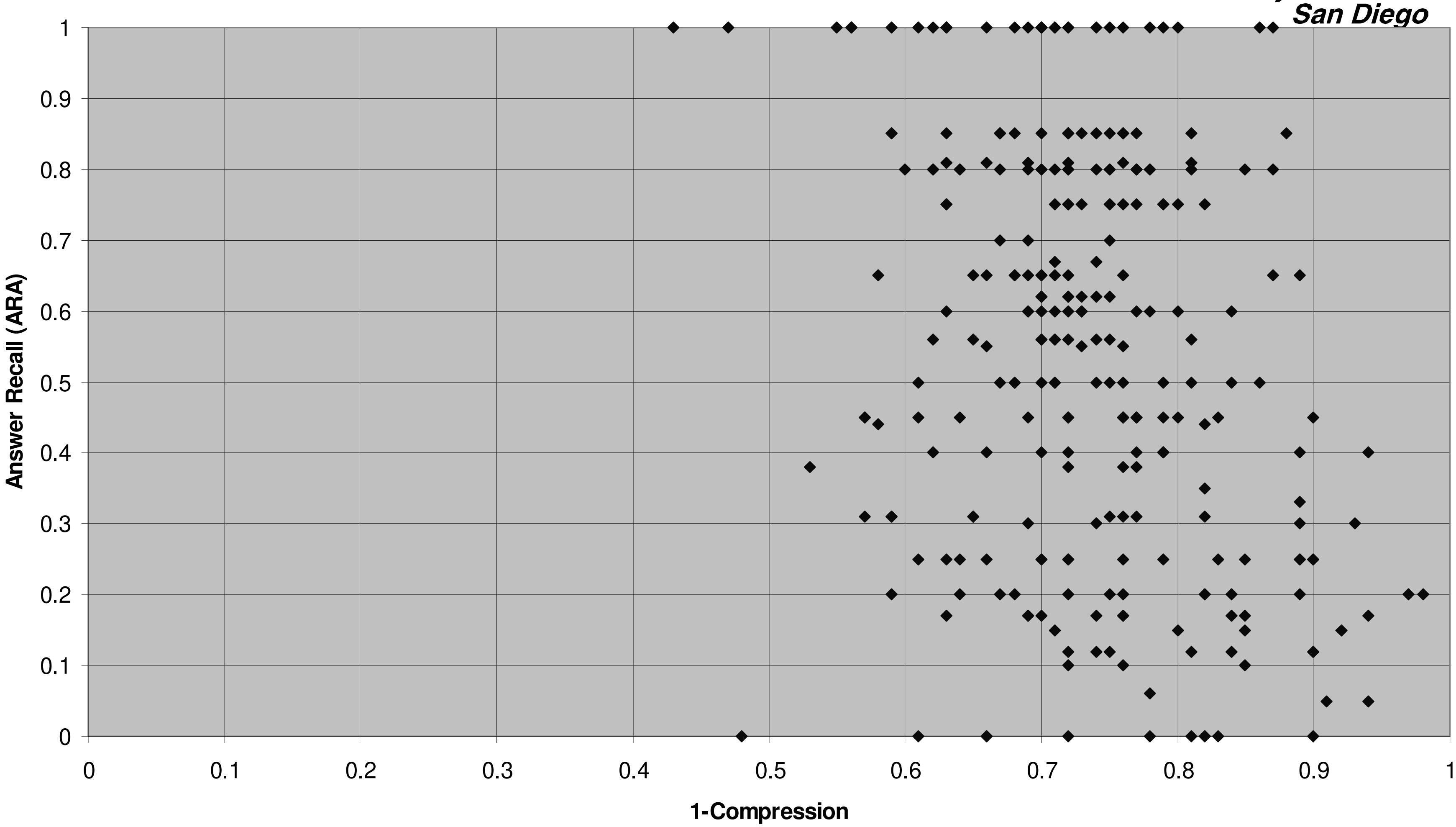
Length of Summaries in Relation to Full Docs for "Secondary Theme" (ST) Topic Prominence Category: Topic 257

SPAWAR
Systems Center San Diego

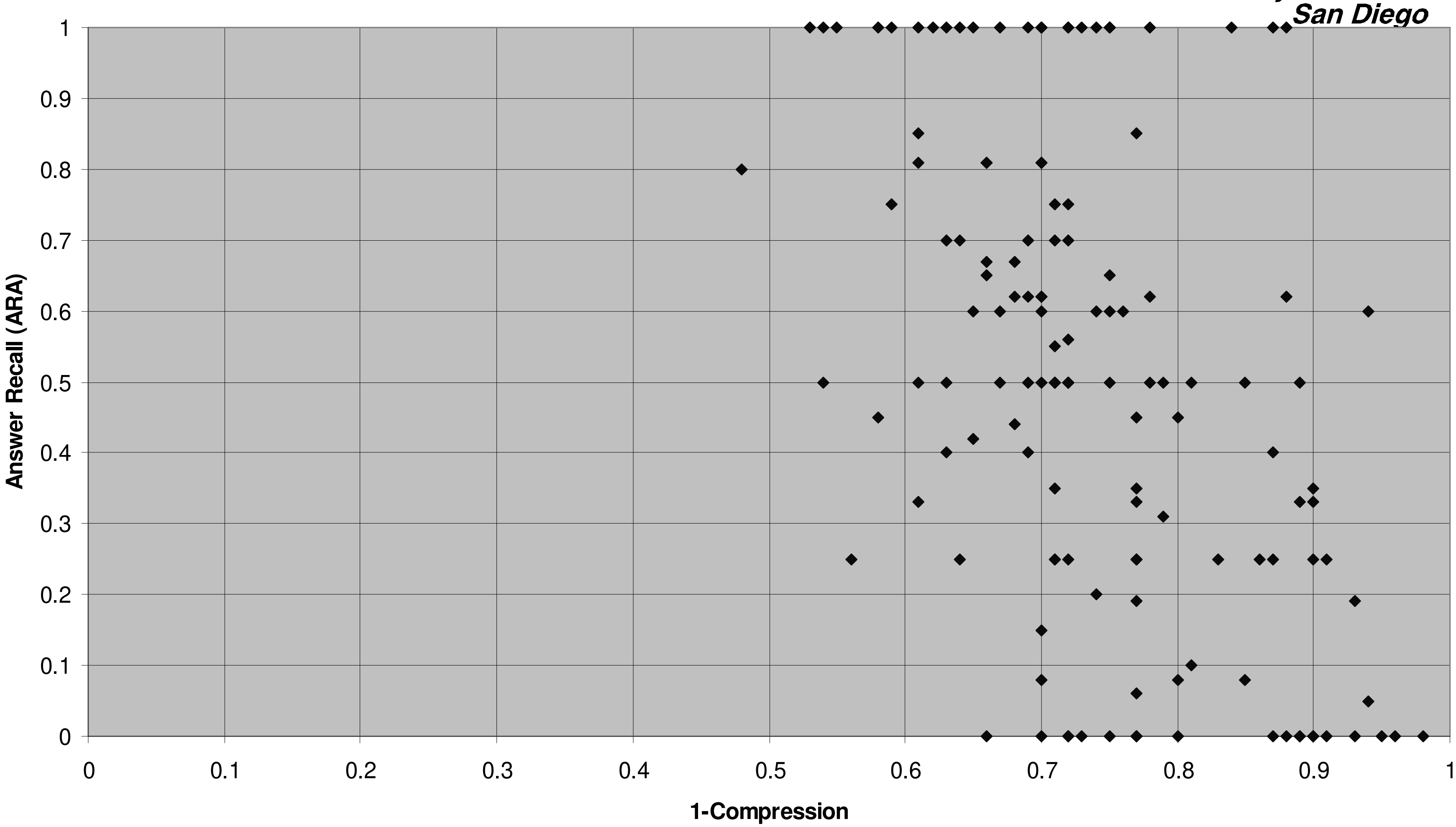


Recall vs Compression for All Docs in "T" Topic Prominence Category

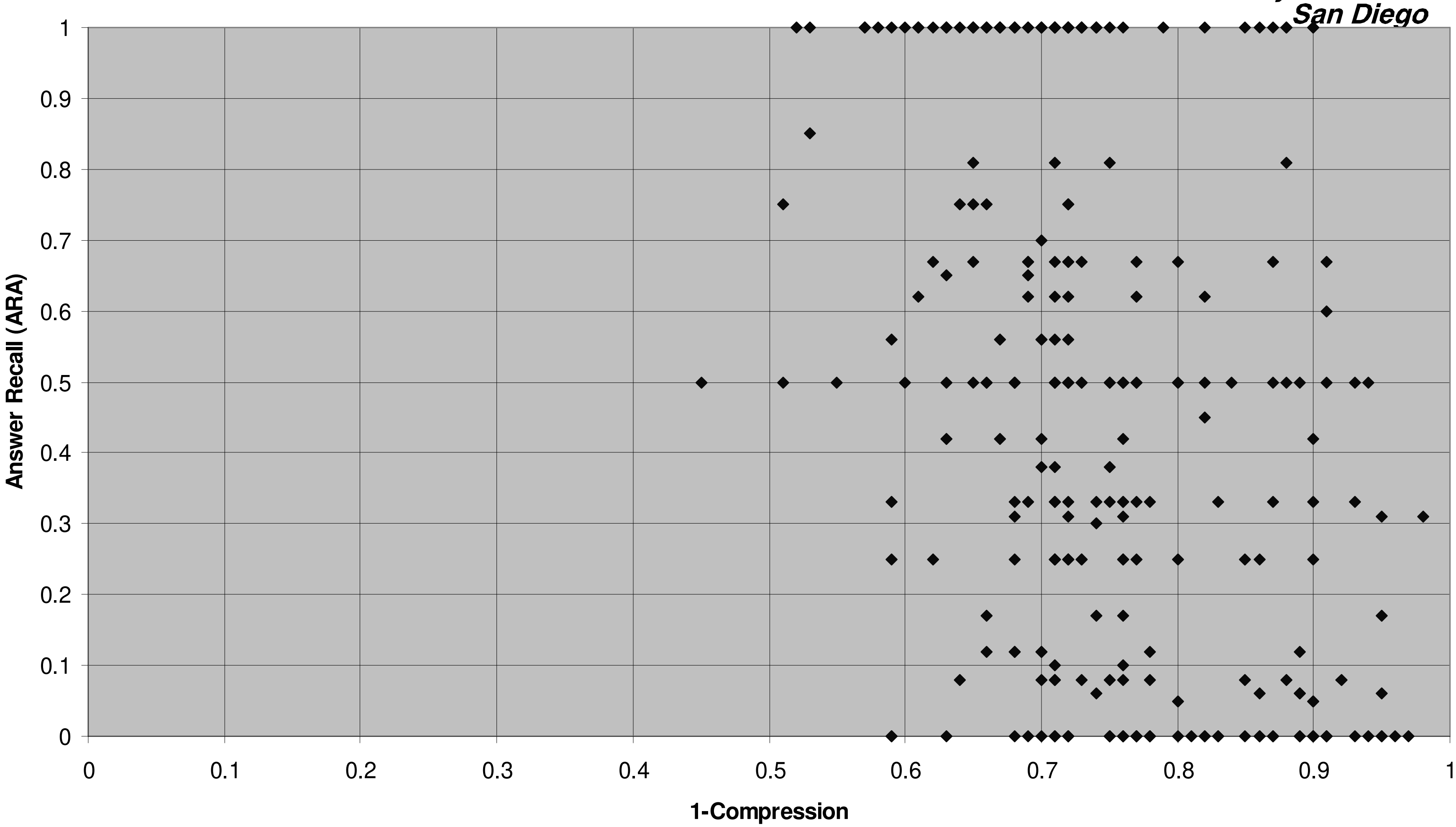
SPAWAR
Systems Center
San Diego



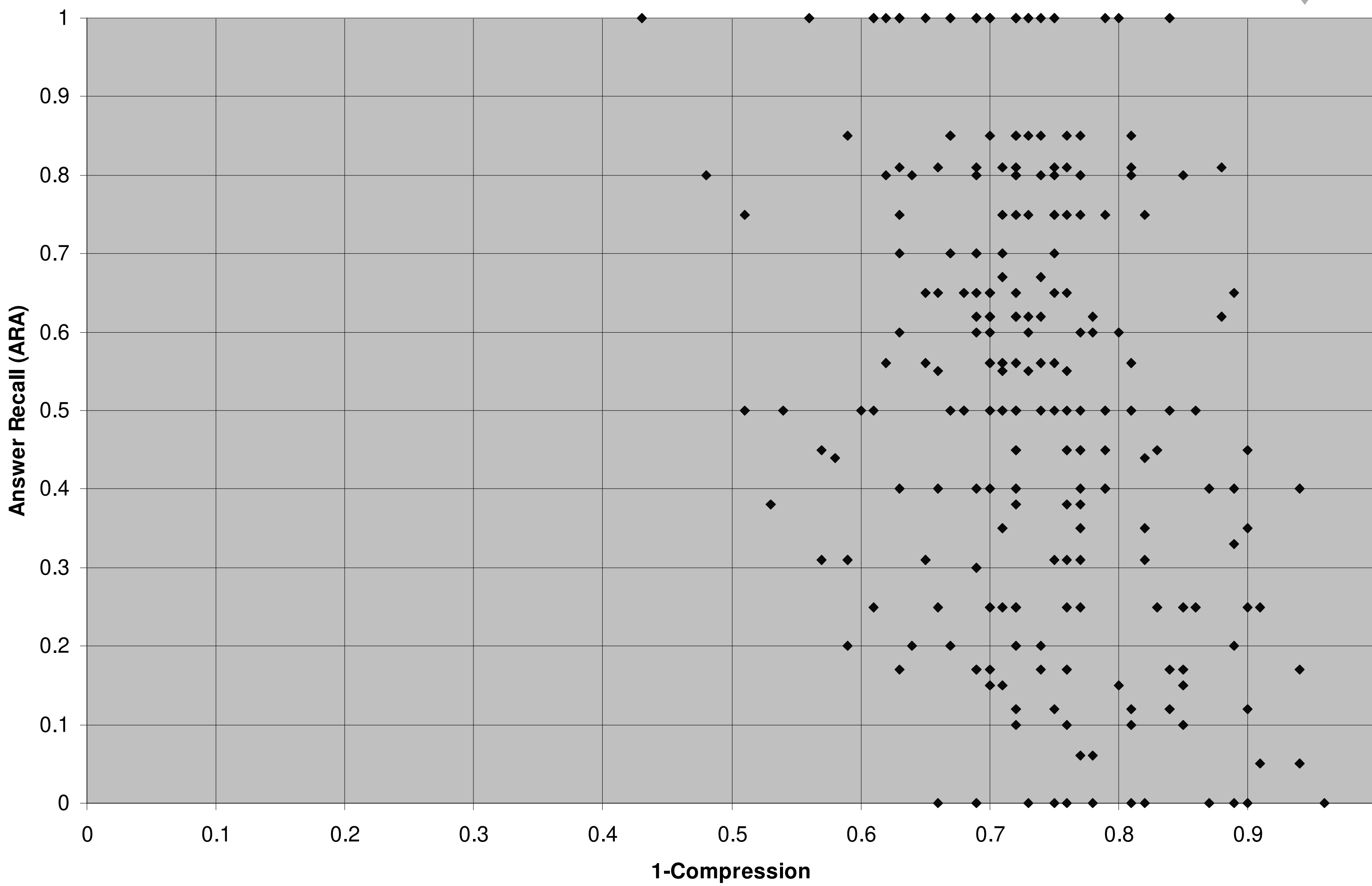
Recall vs Compression for All Docs in "MT" Topic Prominence Category



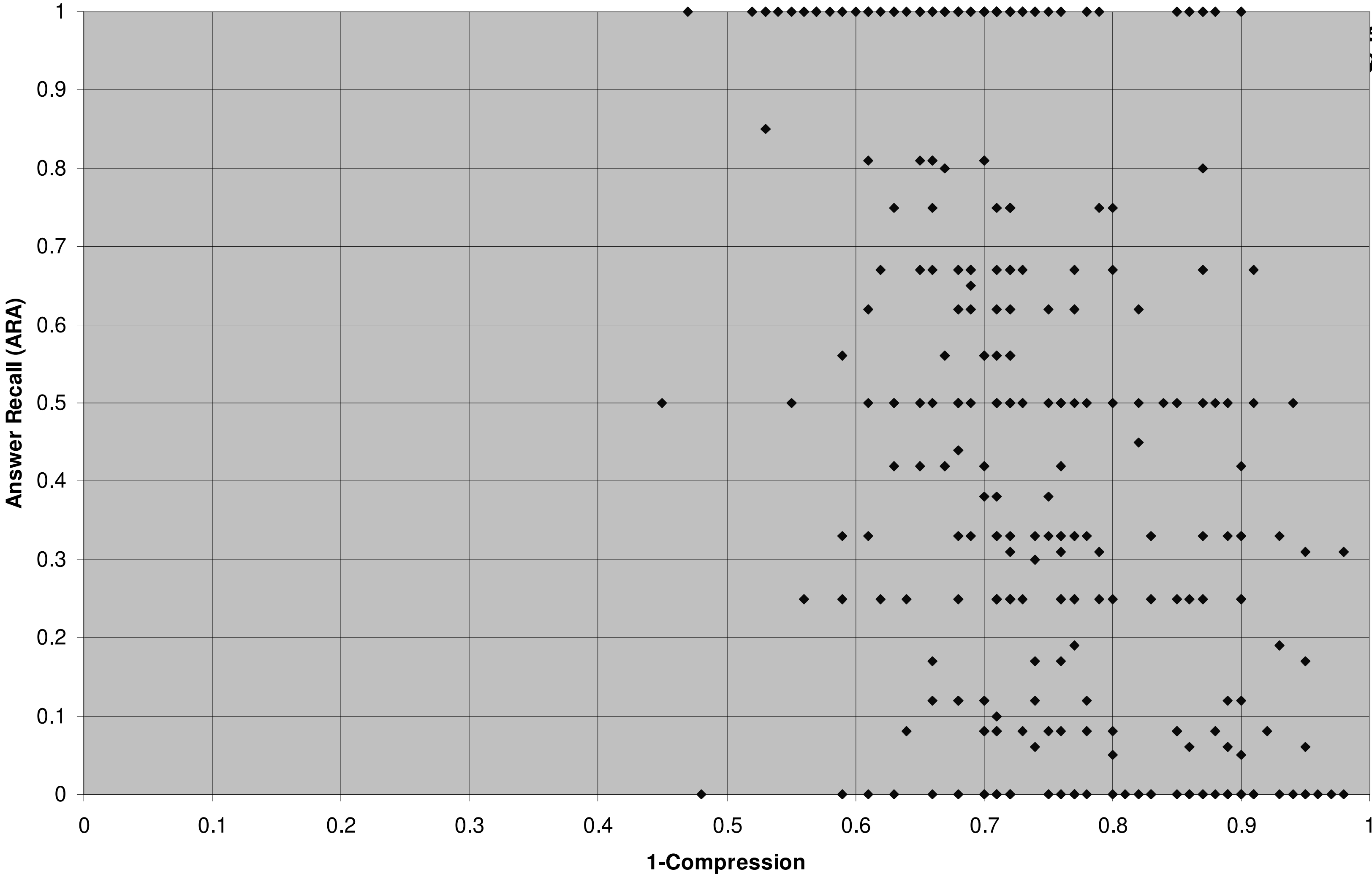
Recall vs Compression for All Docs in "ST" Topic Prominence Category



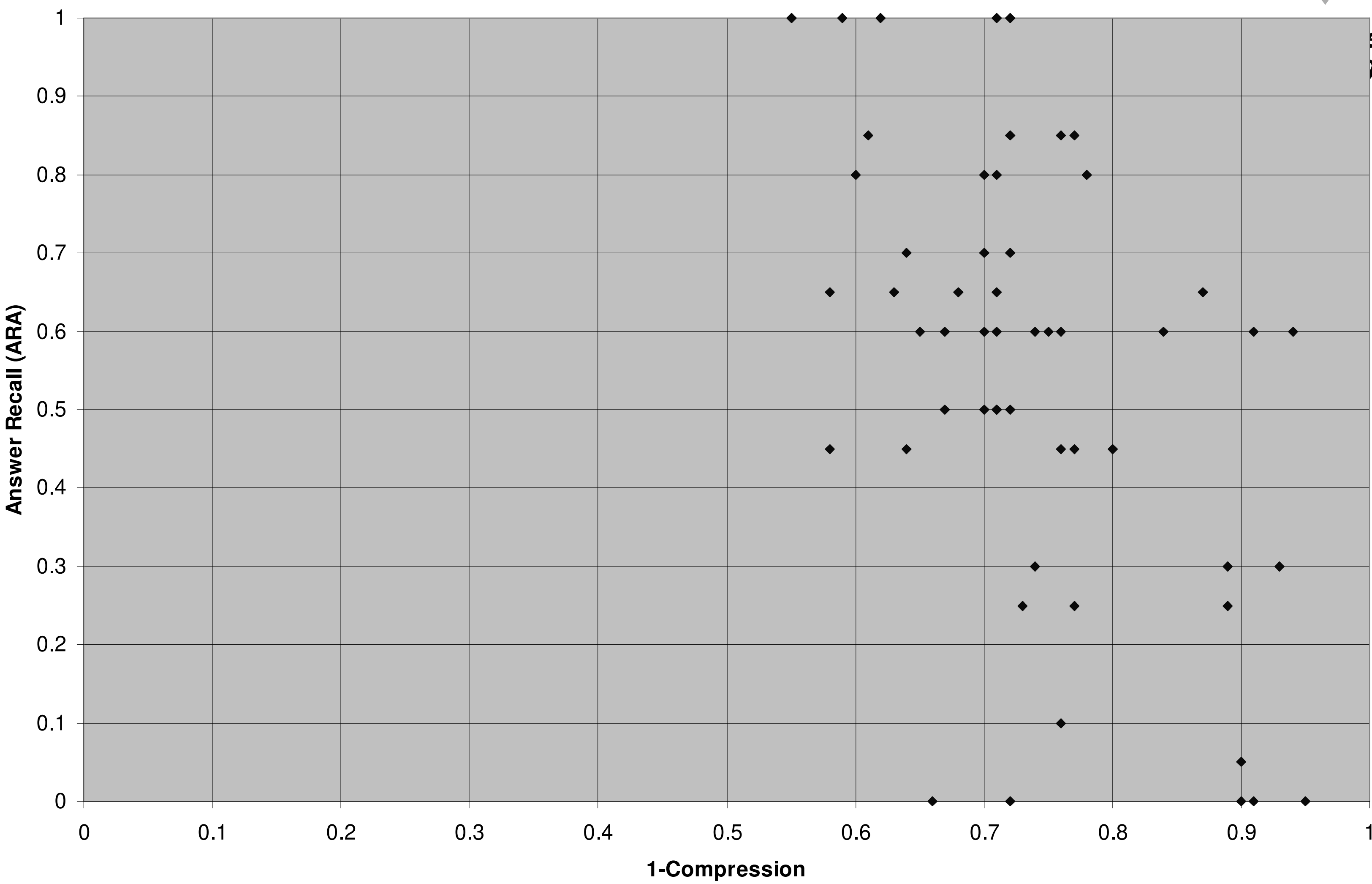
Recall vs Compression for All Docs in "N" (News) Genre Category



Recall vs Compression for All Docs in "F" (Feature) Genre Category



Recall vs Compression for All Docs in "NF" (News/Feature) Genre Category



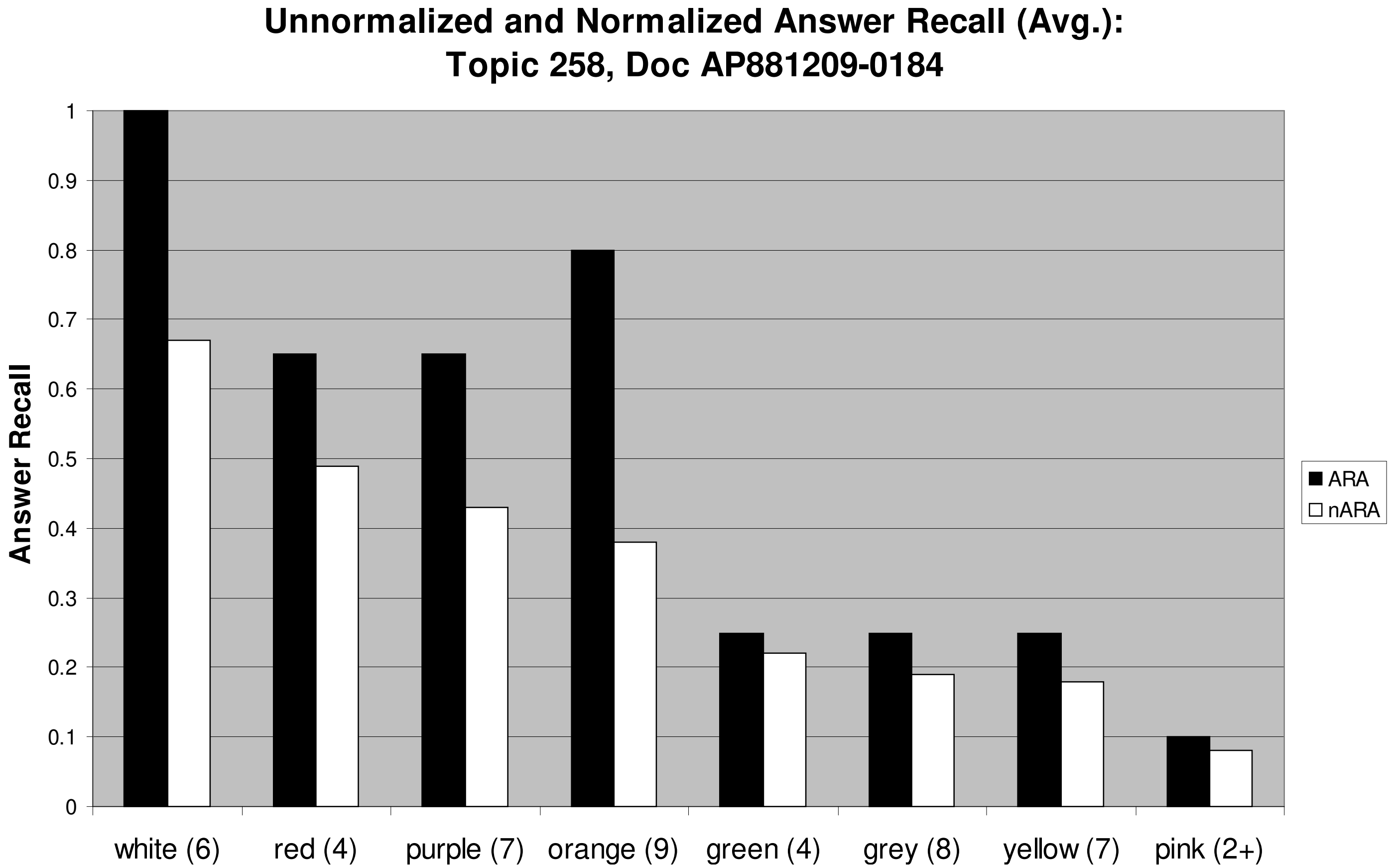
System Performance on Sample Document: Topic 258, Doc AP881209-0184

Systems Center
San Diego

Group performance in terms of scoring decisions made (total n per question = 8)

Score Cat	*Q1:Who*	*Q2:How*	*Q3ToWhom*	*Q4:What*	*Q5:When*
Missing	0	4	4	7	0
Partial	1	0	0	0	6
Correct	7	4	4	1	2

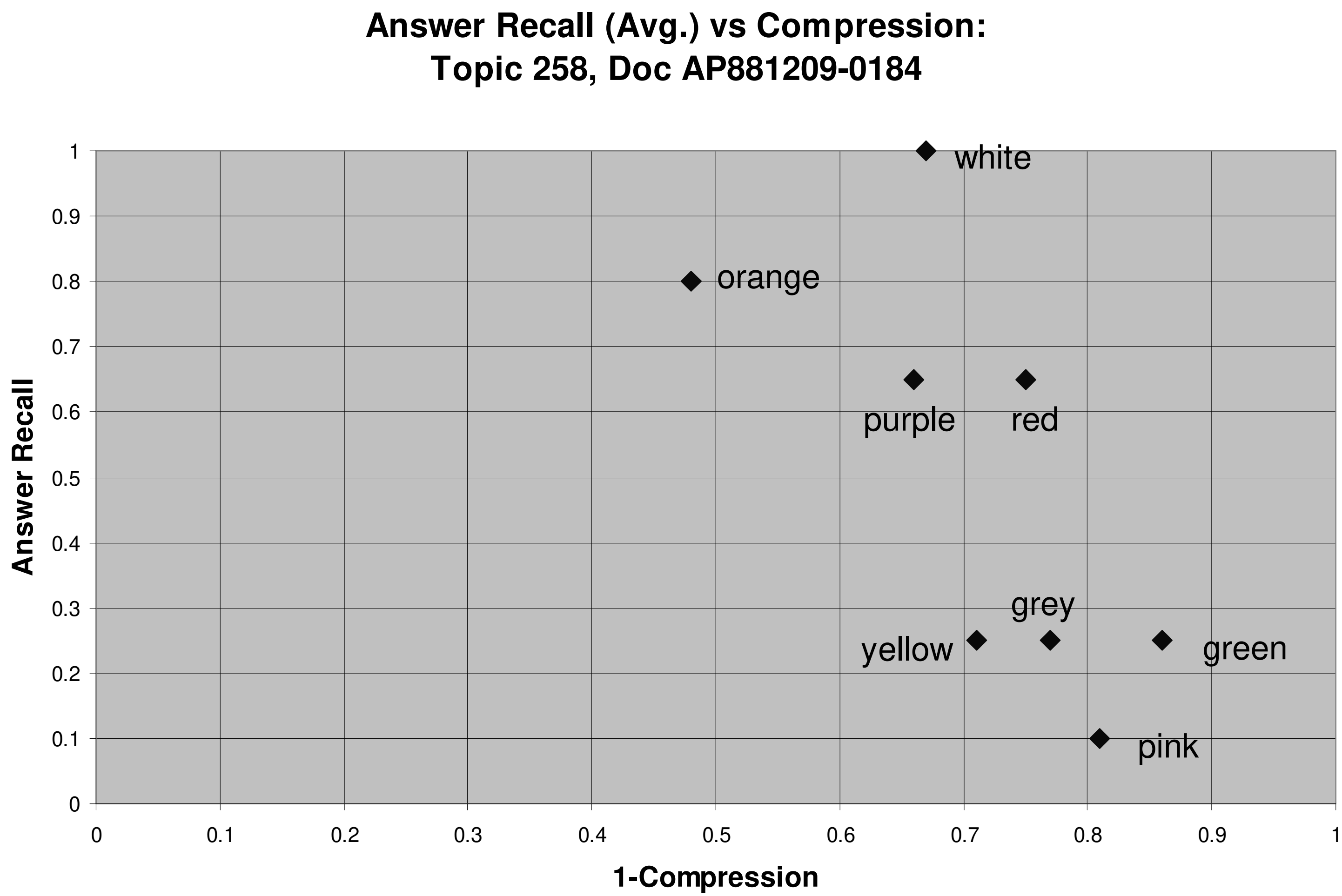
Performance on Sample Doc (Cont'd)



Note: Num in paren's is num *sentences* in summary.

Performance on Sample Doc (Cont'd)

Systems Center San Diego



What Next?
(Your Preferences, Please)

Systems Center
San Diego

- To improve corpus from current eval: Revise keys and model summaries, review scoring
- To enable testing by developers: Sentence-alignment and coarse-grained scoring software
- To support basic R&D: Different types of model summaries (news, biographical, opinion, background, …)
- To support cross-doc summarization R&D: Create cross-doc annotations (e.g., for Robert Morris computer virus incident)