# iFACETSUM: Coreference-based Interactive Faceted Summarization for Multi-Document Exploration

**Eran Hirsch[1], Alon Eirew[1,2*], Ori Shapira[1*], Avi Caciularu[1], Arie Cattan[1],**
**Ori Ernst[1], Ramakanth Pasunuru[3], Hadar Ronen[4], Mohit Bansal[3], Ido Dagan[1]**

[1]Bar-Ilan University    [2]Intel Labs, Israel    [3]UNC Chapel Hill    [4]Peres Academic Center
{hirsch.eran, obspp18}@gmail.com
alon.eirew@intel.com dagan@cs.biu.ac.il

## Abstract

We introduce iFACETSUM,[1] a web application for exploring topical document sets. iFACETSUM integrates interactive summarization together with faceted search, by providing a novel faceted navigation scheme that yields abstractive summaries for the user's selections. This approach offers both a comprehensive overview as well as concise details regarding subtopics of choice. Fine-grained facets are automatically produced based on cross-document coreference pipelines, rendering generic concepts, entities and statements surfacing in the source texts. We analyze the effectiveness of our application through small-scale user studies, which suggest the usefulness of our approach.

## 1 Introduction

An information consumer aspiring to explore a new topic will often be faced with an extensive collection of texts from which to acquire knowledge. Confronted with these texts, the reader would have difficulty determining where to start reading and obtaining details about specific aspects of the topic. Addressing this we present iFACETSUM, illustrated in Figure 1, an *interactive faceted summarization* approach and system for navigating within a large input document-set on a topic. The system initially provides a full high-level overview of the topic at a glance in the form of facets. A user can then dive further into subtopics of interest and obtain concise facet-based summaries, capturing the valuable information of a subtopic.

The challenge of knowledge navigation has been addressed with various solutions, mainly under the umbrella of *exploratory search* (Marchionini, 2006) tasks. For example, in Complex Interactive Question Answering (ciQA) (Kelly and Lin, 2007) and

Conversational QA (Reddy et al., 2019), a user interacts with a QA system in order to meet an information need on the source text(s). Interactive information retrieval (Ingwersen, 1992) and conversational search (Radlinski and Craswell, 2017) refine document retrieval through different means of textual interaction. Both tasks do not offer a preliminary outline of the source documents, and hence expect a user to formulate queries or questions without system guidance. Furthermore, short answers, such as those output in conversational QA, may be insufficient, while lists of relevant textual results, such as in conversational search, may be overwhelming and provoke an inefficient navigation process.

As a midpoint solution, interactive summarization provides an initial summary as an overview of the topic, and the ability to inquire, via suggested or free-text queries, for more information in the form of summary expansions (e.g. Shapira et al., 2021; Avinesh et al., 2018). Here still, the initial summary, along with the suggested queries, do not produce the full high-level picture, and therefore hints only partially at the possible subtopics that the user might want to explore.

iFACETSUM builds upon the interactive summarization scheme, extending it via the effective faceted search approach (Hearst, 2006a) (§2.1), coupled with facet-based abstractive summarization (§3.2). The presented facet values provide a comprehensive overview of the input topic, while the abstractive summaries deliver concise fine-grained information on selected facet values (see Figure 1). Furthermore, since facets are hierarchically updated in accordance to facet-value selections, navigating deeper into subtopics becomes seamless. In terms of backend implementation, facets are automatically derived over the input document set in a novel manner, based on cross-document coreference resolution (Cattan et al., 2021) and proposition alignment (Ernst et al.,

---

**NATIVE AMERICAN CHALLENGES**

treaties x [5]

**Concepts** [1]
- ☑ treaties
- ☐ violations
- ☐ approval
- ☐ program
- ☐ policy
- ☐ authority
- ☐ case
- ☐ problems

Show all

[6]
treaties (6)
compacts (4)
agreements (4)
deals (3)
treaty (3)
deal (2)
settlement (2)
Treaties (2)
transactions (2)
alliance (2)
agreed (1)
pact (1)
treat (1)
enterprises (1)
accord (1)
dealt (1)
compact (1)

**Entities** [2]
- 37
- ☐ Organization: the Cayugas — 4
- ☐ Location: New York — 3
- ☐ Person: von Gernet — 2
- ☐ the country — 2
- ☐ a treaty with them — 2
- ☐ their land — 2
- ☐ tribes — 1
- ☐ the reservation

Show all

**Statements** [3]
- 5 ☐ Clinton was going to the Pine Ridge Reservation for a visit with the Oglala Sioux nation and to participate in a conference on Native American homeownership and economic development. — 2
- 5 ☐ He said the majority faction led by Fish Carrier moved to Buffalo Creek in western New York and tried to sell off the Cayugas lands to private individuals in deals — 2
- 4 ☐ In time…Washington force-fed treaties to the Crow…bunching…inviting in white settlers as models of how to run a store or farm the land. — 2
- 4
- 3
- 3
- 2
- 1

Show all

[7]

| Person | | Location | | Organization | | Miscellaneous | |
|---|---|---|---|---|---|---|---|
| ☐ McCarty | 29 | ☐ New York | 12 | ☐ Makah | 17 | ☐ tribes | 32 |
| ☐ von Gernet | 9 | ☐ Navajo | 6 | ☐ the Cayugas | 12 | ☐ the county | 17 |
| ☐ Garcia | 6 | ☐ the United States | 5 | ☐ Montoya | 8 | ☐ the tribe 's | 16 |
| ☐ Neah Bay | 4 | ☐ New York City | 5 | ☐ Newton | 8 | ☐ the reservation | 15 |
| ☐ Eric | 3 | ☐ Cambridge | 4 | ☐ Pine Ridge | 8 | ☐ the nation | 15 |
| ☐ Bam Bam | 3 | ☐ Idaho | 4 | ☐ Cutler | 6 | ☐ the Onondagas | 15 |

**Summary** [4]                                                                 Original sentences  History

Tribal leaders hope settlement will bring assets they need to upgrade reservation. Law requires tribes to reach compact with state in which reservation lies if it wants to open a casino. California does not allow gambling in the state, which has refused to allow gambling anywhere. Florida, Kansas and Alabama are trying to block the opening of Indian casinos within their borders.

Figure 1: Our iFACETSUM web application over a set of 25 documents about "Native American Challenges". The user gets an overview of the topic as *Concepts* [1], *Entities* [2] and *Statements* [3] facets. The facets are updated in response to the user's choice of the facet-value "treaties" [5]. An abstractive summary is generated for the set of sentences corresponding to the "treaties" semantic cluster [4]. The mentions of a facet-value appear when hovering over its frequency [6]. Clicking "Show all" opens a pop-up with more facet-values. The *Entities* pop-up is categorized into further facets of *Person*, *Location*, *Organization* and *Miscellaneous* [7].

2020), yielding clusters of facet-value mentions (§3). Accordingly, summaries are generated based on the sentences that contain mentions of *all* selected facet-values.

We conduct usability studies on our system, and demonstrate its utility for easy navigation in topical document sets, while enabling deep diving into desired knowledge without losing the context of the exploration process.

We next describe iFACETSUM's interface in §2 and its backend implementation in §3. This is followed by the description and results of our usability investigations in §4, an overview of related work in §5, and finally conclusions and suggestions for future work in §6.

## 2 iFACETSUM Interface

iFACETSUM is a web application for exploring a document-set on a topic, shown in Figure 1. It generally consists of the *faceted navigation* component (top of figure, described in §2.1), and the facet-based *summary* component (bottom of figure, §2.2). The former rests upon a faceted-navigation panel that provides orientation on the source topic, while the latter supplies the user with key information about selected facet-values. This flow facilitates guided exploration, over the full scope of the topical information and within subtopics of interest.

### 2.1 Faceted Navigation

Faceted search is a technique used to provide more effective information-seeking support (Tunkelang, 2009), by allowing users to narrow down results based on rich attributes. A *facet* describes an attribute type, and *facet-terms* or *facet-values* represent attribute values. iFACETSUM's facets are formed using techniques that identify recurring mentions of sub-sentential units in texts, as explained further in §3.1.

The faceted navigation component is laid-out to the user in the form of three general facets (Figure 1, [1], [2] and [3]): (1) *Generic Concepts* facet, e.g., "poverty" and "treaties". (2) *Entities* facet, containing values such as e.g., "Clinton" as a person or "Nebraska" as a location. (3) *Statements* facet, which lists specific statements mentioned several times, such as "Nebraska does not allow casino gambling".

In our data scheme, each facet-value encapsulates a cluster of mentions that semantically refer to a common concept, entity or statement, and, as such, may be lexically diverse (e.g., the "case" concept associates with mentions of "lawsuit", "fight", "battle", "debate"). A *facet-value sentence-set* is defined as the set of sentences pertaining to all of a facet-value's mentions. The *facet-value label* is the facet-value name presented to the user, and is chosen to be the most frequent lexical type in the mention cluster corresponding to that facet-value.

The values under each facet are ordered by their frequencies (number of mentions) in the source document set, as an indication for level of salience. A facet-value is shown with its corresponding frequency, and its various mention forms are revealed by hovering over the frequency meter (e.g., depicted in [6], the cluster "treaties" includes mentions of "agreements", "deals", etc.). Only a few of the top facet-values are shown under each facet, while clicking *Show all* expands the facet in full, in a pop-up. The pop-up for the *Entities* facet partitions the facet-values to particular sub-facets: *Person*, *Location*, *Organization* and *Miscellaneous* ([7]).

By clicking a facet-value, the system generates a summary of its sentence-set. Additionally, the facets update to include only values appearing in that sentence-set. The updated facet view thus gives an overview which is fine-grained for the selected subtopic, while iteratively selecting additional facet-values supports diving deeper into it. When additional facets are gradually selected, a summary is generated over the intersection of the sentence-sets of all selected facets. Any of the selected facet-values can be canceled out, whereby the system updates accordingly.

## 2.2 Facet-based Summarization

Upon a change in selection of facet-values, the system provides the user with targeted information via an abstractive summary of the selections' sentence-set ([4]). As more facet-values are selected, the generated summary is based on the intersection of the sentence sets of all selected facets, becoming more specific. The user can further view the complete set of source sentences used to generate the summary, and those sentences' full documents (Figures 3 and 4 in Appendix). Additionally, clicking "History" shows all previously generated summaries (Figure 5 in Appendix).

## 3 Backend Algorithms

As portrayed in §2, iFACETSUM supports two central features: presenting a faceted navigation panel and generating a summary around selected facet-values. We next describe how facet-values are generated using CD coreference resolution (§3.1), and how we apply abstractive summarization, based on a facet-value selection (§3.2). Figure 2 illustrates the entire process.

## 3.1 Coreference-based Facet Formation

As described in §2.1, there are three main facets. *Concepts* and *Entities* are extracted using cross-document (CD) coreference resolution pipelines, while *Statements* via a proposition alignment pipeline, described next.[2]

**Concepts.** We found that identifying and grouping together significant co-occurring events within the source document collection helps to expose and emphasize the notable concepts in the topic. To that end, we employ CD **event coreference resolution** which detects these concepts.

CD coreference resolution (Lee et al., 2012) clusters text mentions that refer to the same event or entity across multiple documents. Presently, the Cross-Document Language Model (CDLM) (Caciularu et al., 2021) is the state-of-the-art for CD coreference resolution. This model is pretrained on multiple related documents via cross-document masking, encouraging the model to learn cross-document and long-range relationships. Specifically, we employ the CDLM version fine-tuned for coreference on the ECB+ corpus (Cybulska and Vossen, 2014). This model does not include a mention detection component, but rather expects relevant mentions to be marked within the input texts. We therefore leverage the mention detection ability of the model by Cattan et al. (2021).

Once we have obtained the coreference clusters from CDLM, events whose mentions are predominantly verbs are filtered out,[3] since those usually present specific actions that tend to be less informative compared to nominal types that refer to more generic events (e.g., "said", "found" "increase" compared to "unemployment", "poverty", "crash").

CD event coreference resolution separates specific event instances, hence differentiating between clusters of similar event types with different arguments (e.g., "unemployment" in Navajo vs. "unemployment" in Cayuga). Since generic event types, like "unemployment", are more suitable as facet-values, clusters with the same label (most frequent mention) are merged. Each such merged clusters then constitutes a single facet-value, to be presented to the user as described in §2.1.[4]

---

[2]Facet extraction runs in a pre-processing step, since it is not fast enough for real-time latency (see Appendix A.2).

[3] Using spaCy (Honnibal et al., 2020).

[4]We observed that the CD event coreference model has a tendency to wrongly collapse events of the same type, effec-
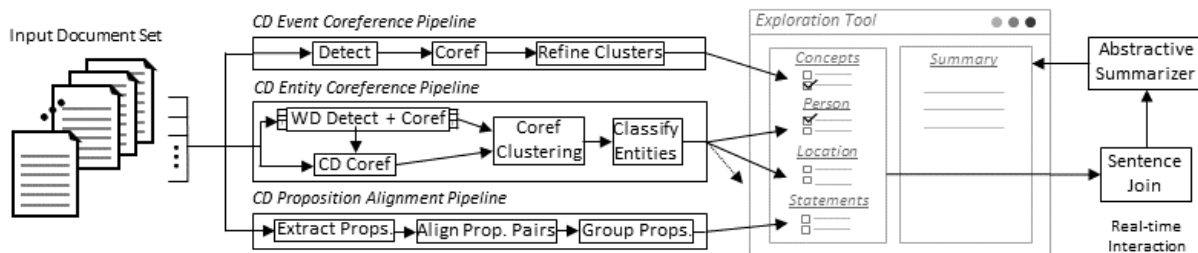
Figure 2: The iFACETSUM architecture. CD = cross-document, WD = within-document.

**Entities.** The *Entities* facet-values help the user focus on entities such as people (e.g., "Clinton"), locations (e.g., "New York"), organizations (e.g., "FBI") and others (e.g., "the casino"). We created a separate pipeline for CD **entity coreference resolution**, since we observed subpar performance when applying the above CD coreference pipeline for entity coreference.[5]

Unlike event coreference, mostly studied in the CD setting, entity coreference has recently seen impressive progress in the *within*-document (WD) setting (Wu et al., 2020; Joshi et al., 2020). Hence, we leverage WD entity coreference in our entity recognition pipeline, which comprises three main steps. (1) We use SpanBERT[6] (Joshi et al., 2020), a state-of-the-art transformer-based LM for WD entity coreference resolution, to detect and cluster coreferring entity mentions within each separate document. (2) The entity mentions detected in the first step are marked as input for a CD entity coreference reolution model. To overcome ECB+ entity scarcity referred earlier, we use an alternative model that is trained on the WEC-Eng dataset (Eirew et al., 2021).[7] (3) Finally, we apply agglomerative clustering to combine the coreference clusters from steps 1 and 2 (WD and CD), and produce the overall entity coreference clusters (details in Appendix A.2).

Once all entity coreference clusters are extracted, we bin them into more specific categories ("Person", "Location" and "Organization"), as portrayed in §2.1, by invoking a Named Entity Recognition (NER) model.[3] A facet-value cluster is tagged with the majority NER label of the mentions in the cluster, among Person, Organization and Location. If

no NER label is assigned to a cluster, it is tagged as "Miscellaneous" (more details in Appendix A.2).

**Statements.** Key statements benefit a user by presenting information about specific facts. To generate these statements, we group together coreferring propositions (rather than words) that describe the same fact within the source documents, as seen in §2.1.

Following Ernst et al. (2020), our pipeline consists of three steps. (1) Proposition candidates are extracted with OpenIE (Stanovsky et al., 2018). (2) Pairs of propositions expressing the same statement are matched using the SuperPAL model (Ernst et al., 2020), considering proposition pairs whose alignment score is above 0.5 as matched. (3) A propositions graph is created by connecting pairs of nodes that represent similar propositions, and proposition clusters are matched for the connected components in the graph (more details in Appendix A.2).

### 3.2 Abstractive Facet Summarization

In the standard summarization setting, a system receives a single or multiple documents as input, as well as a query in the query-focused task. In our case, the input is a set of sentences that have one or more selected facet-values in common, effectively providing a *multi-facet* summary. Given the set of sentences that correspond to the facet-value selection(s), these sentences are concatenated, ordered by their position in their source document (more details in Appendix A.2). This text is then given as input to BART (Lewis et al., 2020), a denoising sequence-to-sequence model fine-tuned on the single-document abstractive summarization task.[8]

iFACETSUM presents abstractive rather than extractive summaries due to their enhanced readability, particularly when summarizing a set of related sentences. This choice follows prior work, which

---

tively aiding our concept formation.

[5]This is in line with previous work (Cattan et al., 2021) which points out that the ECB+ dataset only considers entities that are arguments of event mentions, which is non-exhaustive.

[6]Using AllenNLP (Gardner et al., 2018).

[7]Fine-tuning CDLM on WEC-Eng is computationally infeasible, and therefore we use the model by Eirew et al. (2021).

[8]We use the huggingface model from https://hugg ingface.co/facebook/bart-large-cnn.

showed that fusing sentences with shared points of coreference potentially facilitates coherence of abstractive summaries (Lebanoff et al., 2020). Indeed, in an internal manual assessment of 30 random individual summaries produced by iFACETSUM, with 5 readability measures (Dang, 2006), testers found overall that the summaries are highly readable. To verify that factuality is not compromised, an additional inspection found that these summaries were also factually consistent to the input text, with 28 out of 30 sampled sentences marked as consistent. See Appendix B.3 for scores and more details on these assessments.

## 4 System Experiments

iFACETSUM aims to provide an effective means of information seeking in scenarios that require learning or investigating a new topic (Marchionini, 2006). To that end, we tested this goal through two small-scale experiments with human subjects, as a preliminary examination of the system. In the first experiment, we conducted a pilot usability study to inspect whether users felt they were able to satisfactorily complete an information seeking task using our system. In the second, we examined whether iFACETSUM is preferred over a standard document-search system to complete the exploration task.

### 4.1 Usability Study

**Setup.** The purpose of this experiment was to get general feedback, from human subjects, on the usability of the system, following established usability study methodologies (Nielsen, 1994). To simulate a realistic use case of topic exploration, we instructed participants to use the system in order to prepare a draft review, given an informational goal, that a reporter could then use to write a report on the topic. We prepared guiding story-lines (Appendix B, Table 1), as informational goals, for two topics from the DUC 2006 MDS dataset (NIST, 2005). To analyze iFACETSUM in different exploratory situations, one topic is broad with higher information variability across the articles ("Native American Challenges"), while the other is more focused on a specific event ("EgyptAir Crash").

In this pilot usability study, six participants[9] explored both topics in random order. During system usage we observed the users' activity, via a "think

---

[9]The discount usability testing principle contends that six evaluators are sufficient for prototype evaluation (Nielsen, 1993).

aloud" technique (Van Someren et al., 1994), to obtain user remarks. After exploring a topic, a user rated, from 1 to 5, the usefulness of different aspects of each component in the interface. After both topics, a System Usability Scale (SUS) questionnaire (Brooke, 1995) was filled, to assess global usability of the system (overall score from 0 to 100). Further details are available in Appendix B.1.

**Results.** The average SUS score over the 6 participants is 82.9, where 80.3 is considered "excellent" (UIUX-Trend, 2021). From the average component ratings over the 12 sessions, users expressed their satisfaction with the facet view's and summaries' quality for the use of the tasks. The overall facets quality received a score of 4.3 (SD=0.7), summary coherence 4.7 (SD=0.5), summary informativeness 4.2 (SD=1.1), summary non-redundancy 3.8 (SD=1.0), and summary length 4.3 (SD=0.9). General feedback and issues raised by participants are available in Appendix B.1. Overall, participants were pleased with their experience and some voiced their desire to use the tool right away for current event issues, like COVID-19 vaccination.

As expected, users noticed a difference between the two topics, and mentioned that they preferred the *Concepts* facet for "Native American Challenges", while preferring the *Entities* facet for "EgyptAir Crash". Users found the *Statements* facet-values rather lengthy and less useful, and at times considered it a substitute for summarizing the topic. Future improvements of the system may include considering alternative uses of the aligned statements, like linking specific fact mentions across documents.

### 4.2 Comparative Analysis

To further investigate whether iFACETSUM is an effective tool for exploring a new topic, we conducted a small-scale comparison with a search tool, which roughly simulates common means for learning about a new topic. We asked four new experimentees to carry out the exploration task described in §4.1, once with our system on one topic, and once with the search tool on the other topic (in different orders). The search tool used was DocFetcher,[10] an open source desktop search application, which indexes the given files, enables searching documents with queries, and highlights

---

[10]http://docfetcher.sourceforge.net

query terms within retrieved documents. The participants finished their assignment with iFACETSUM slightly faster than with the search tool. More importantly, they conveyed their satisfaction of using iFACETSUM as a tool for navigating through multiple texts, and learning about a new topic. The participants filled a questionnaire, rating each question on a scale of 1 (DocFetcher is preferred) to 7 (iFACETSUM is preferred). The questions included: (1) Which system was *easier to use* in order to get the desired result? (Avg=5.5, SD=1.73); (2) With which system was it easier for you to *get an overview* of the topic? (Avg=5, SD=2.3); (3) With which system was it easier for you to get *detailed information* about a subtopic of interest? (Avg=5.25, SD=0.9); (4) If you had to learn about or explore a new topic, *which system would you choose*? (Avg=5.25, SD=0.95). Overall, participants favored iFACETSUM in all questions, preferring it for future use (details in Appendix B.2).

## 5 Related Work

Attaining information of interest from large document sets has been approached with different techniques. A vast amount of research has been conducted on *multi-document summarization*, as a method for presenting the central aspects of a target set of texts (e.g. Barzilay et al., 1999; Haghighi and Vanderwende, 2009; Bing et al., 2015; Yasunaga et al., 2017), where *query-focused summarization* (Dang, 2005) biases the output summary around a given query (e.g. Daumé III and Marcu, 2006; Baumel et al., 2018; Xu and Lapata, 2020).

Recognizing the need for dynamically acquiring a broader or deeper scope of the source texts, *exploratory search* (Marchionini, 2006; White and Roth, 2009) was coined as an umbrella term for allowing more dynamic interactive exploration of information. Adapting the summarization paradigm to the exploratory setting, *interactive summarization* enables a user to refine or expand on a summary via different modes of interaction. For example, Shapira et al. (2021), Avinesh et al. (2018) and Baumel et al. (2014) provide a limited (or no) initial summary on the document set, and support iterative interaction, via queries or preference highlights, to update the summary. However, the succinct initial summary, possibly accompanied by few suggested queries, do not display the *full* scope of the source texts, which limits the user's perception of the many available sub-topics to learn more about.

On the other hand, other exploratory search approaches *do* provide a more elaborate overview of the source data through sophisticated dashboards or facets of extracted information or metadata (e.g. O'Connor et al., 2010; Koren et al., 2008; Hope et al., 2020). Indeed, faceted navigation (Hearst, 2006a; Tunkelang, 2009) is an effective instrument for navigating within a large data source (Hearst et al., 2002; Ruotsalo et al., 2020). While most faceted search systems generate facets from semi- or fully-structured data, as prominently encountered in e-commerce websites and in research (Hearst, 2006b; Ben-Yitzhak et al., 2008), some works generate facet hierarchies from unstructured open-domain texts. For example, from product reviews, Ly et al. (2011) extract product aspects and present several summaries, each focused on a single aspect as a "facet", in a form of single-level faceted search. Hope et al. (2020) devise facet-values from scientific articles by eliciting unstructured textual information (topics, entities) from the articles and their structured metadata (e.g article authors). Although these search tools offer a more comprehensive overview of the source data, they either present raw-text search results or do not allow thorough navigation.

iFACETSUM fully integrates dynamic multi-level faceted navigation into interactive multi-document summarization. The facets serve as an efficient means of grasping the topic, and render an intuitive medium for navigating through the information. The abstractive summaries generated at real-time expose concise details for any combination of sub-topics of choice. Furthermore, we innovatively employ coreference resolution and proposition alignment to generate fine-grained open-domain facets.

## 6 Conclusion and Future Work

In this paper, we presented iFACETSUM, a novel text exploration approach and tool over large document sets, which incorporates faceted search into interactive summarization. Its faceted navigation design provides a user with an overview of the topic and the ability to gradually investigate subtopics of interest, communicating concise information via multi-facet abstractive summarization. Fine-grained facet-values are generated from the source texts based on cross-document coreference pipelines. Small-scale user studies suggest the utility of our approach for exploring a new topic from

multiple documents.

Future work may speed up the coreference-based facet extraction pipeline, allowing for real-time processing of ad-hoc document sets, and may investigate further methods for facet generation. Additional search techniques might be integrated into the exploration scheme, including free text searching as raised in the user study. It would also be appealing to try adapting the system to domains other than news, such as the medical or scientific domains, for which exploration tools would be very useful. Such adaptations would depend on the portability of the underlying technologies of cross-document coreference resolution and proposition alignment. Finally, future work may explore additional ways of leveraging the power of recent proposition alignment methods.

## 7 Ethical Considerations

**Usability Study.** We conducted the usability study (§4.1) over Zoom sessions (`https://zoom.us/`), and carried out the "think aloud" technique through screen sharing and a with an open camera. Participants volunteered to take part in the study, and took about 45 minutes of each of their time. An *informed consent* form was signed by the participant before each study.

The comparative study included four NLP doctoral students from our lab who volunteered for the experiment. The summary readability and factual consistency assessments were done by two authors of this paper.

**Computation.** We ran the three pre-processing pipelines mentioned in §3.1 on 2 to 4 GPUs, where each pipeline ran from a few minutes to 10 hours per topic (25 news articles). Six such topics were prepared for the demo applications. (more details in Appendix A.2).

The summarization model runs in real-time (per user interaction) over a CPU in less than 3 seconds per summary. Summaries are cached to refrain from recomputing summaries for repeated queries.

**Dataset.** The DUC 2006 data was acquired according to the required NIST guidelines (`duc.nist.gov`).

**Multilingualism.** All models used within the components of iFACETSUM were trained on English data, thus making the system compatible for English only. Supporting other languages requires replacing the contained models to ones compliant to the desired languages.

## References

PVS Avinesh, Benjamin Hättasch, Orkan Ozyurt, Carsten Binnig, and Christian M Meyer. 2018. Sherlock: A System for Interactive Summarization of Large Text Collections. *Proceedings of the VLDB Endowment*, 11(12).

Shany Barhom, Vered Shwartz, Alon Eirew, Michael Bugert, Nils Reimers, and Ido Dagan. 2019. Revisiting joint modeling of cross-document entity and event coreference resolution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4179–4189, Florence, Italy. Association for Computational Linguistics.

Regina Barzilay, Kathleen R. McKeown, and Michael Elhadad. 1999. Information Fusion in the Context of Multi-Document Summarization. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 550–557, College Park, Maryland, USA. Association for Computational Linguistics.

Tal Baumel, Raphael Cohen, and Michael Elhadad. 2014. Query-Chain Focused Summarization. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 913–922, Baltimore, Maryland. Association for Computational Linguistics.

Tal Baumel, Matan Eyal, and Michael Elhadad. 2018. Query Focused Abstractive Summarization: Incorporating Query Relevance, Multi-Document Coverage, and Summary Length Constraints into seq2seq Models.

Ori Ben-Yitzhak, Nadav Golbandi, Nadav Har'El, Ronny Lempel, Andreas Neumann, Shila Ofek-Koifman, Dafna Sheinwald, Eugene Shekita, Benjamin Sznajder, and Sivan Yogev. 2008. Beyond Basic Faceted Search. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, page 33–44, New York, NY, USA. Association for Computing Machinery.

Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca Passonneau. 2015. Abstractive Multi-Document Summarization via Phrase Selection and

Merging. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1587–1597, Beijing, China. Association for Computational Linguistics.

John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189.

Avi Caciularu, Arman Cohan, Iz Beltagy, Matthew E. Peters, Arie Cattan, and Ido Dagan. 2021. CDLM: Cross-Document Language Modeling. In *Findings of the Association for Computational Linguistics: EMNLP 2021*.

Arie Cattan, Alon Eirew, Gabriel Stanovsky, Mandar Joshi, and Ido Dagan. 2021. Cross-document coreference resolution over predicted mentions.

Agata Cybulska and Piek Vossen. 2014. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4545–4552, Reykjavik, Iceland. European Language Resources Association (ELRA).

Hoa Trang Dang. 2005. Overview of DUC 2005. In *Proceedings of the document understanding conference*, volume 2005, pages 1–12.

Hoa Trang Dang. 2006. Overview of DUC 2006. In *Document Understanding Conference*.

Hal Daumé III and Daniel Marcu. 2006. Bayesian Query-Focused Summarization. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 305–312, Sydney, Australia. Association for Computational Linguistics.

Alon Eirew, Arie Cattan, and Ido Dagan. 2021. WEC: Deriving a large-scale cross-document event coreference dataset from Wikipedia. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2498–2510, Online. Association for Computational Linguistics.

Ori Ernst, Ori Shapira, Ramakanth Pasunuru, Michael Lepioshkin, Jacob Goldberger, Mohit Bansal, and Ido Dagan. 2020. SuperPAL: Supervised Proposition ALignment for Multi-Document Summarization and Derivative Sub-Tasks.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. AllenNLP: A deep semantic natural language processing platform. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 1–6, Melbourne, Australia. Association for Computational Linguistics.

Aria Haghighi and Lucy Vanderwende. 2009. Exploring Content Models for Multi-Document Summarization. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 362–370, Boulder, Colorado. Association for Computational Linguistics.

Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee. 2002. Finding the flow in web site search. *Communications of the ACM*, 45(9):42–49.

Marti A. Hearst. 2006a. Clustering Versus Faceted Categories for Information Exploration. *Communications of the ACM*, 49(4):59–61.

Marti A. Hearst. 2006b. Design recommendations for hierarchical faceted search interfaces. In *ACM SIGIR workshop on faceted search*, pages 1–5.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Tom Hope, Jason Portenoy, Kishore Vasan, Jonathan Borchardt, Eric Horvitz, Daniel Weld, Marti Hearst, and Jevin West. 2020. SciSight: Combining faceted navigation and research group detection for COVID-19 exploratory scientific search. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 135–143, Online. Association for Computational Linguistics.

Peter Ingwersen. 1992. *Information retrieval interaction*, volume 246. Taylor Graham London.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Diane Kelly and Jimmy Lin. 2007. Overview of the TREC 2006 ciQA task. In *ACM SIGIR Forum*, volume 41, pages 107–116. ACM New York, NY, USA.

Kian Kenyon-Dean, Jackie Chi Kit Cheung, and Doina Precup. 2018. Resolving Event Coreference with Supervised Representation Learning and Clustering-Oriented Regularization. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 1–10, New Orleans, Louisiana. Association for Computational Linguistics.

Jonathan Koren, Yi Zhang, and Xue Liu. 2008. Personalized Interactive Faceted Search. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, page 477–486, New York, NY, USA. Association for Computing Machinery.

Logan Lebanoff, John Muchovej, Franck Dernoncourt, Doo Soon Kim, Lidan Wang, Walter Chang, and Fei Liu. 2020. Understanding Points of Correspondence between Sentences for Abstractive Summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 191–198, Online. Association for Computational Linguistics.

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, and Dan Jurafsky. 2012. Joint Entity and Event Coreference Resolution across Documents. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 489–500, Jeju Island, Korea. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. {BART}: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Duy Khang Ly, Kazunari Sugiyama, Ziheng Lin, and Min-Yen Kan. 2011. Product Review Summarization based on Facet Identification and Sentence Clustering.

Gary Marchionini. 2006. Exploratory Search: from Finding to Understanding. *Communications of the ACM*, 49(4):41–46.

Yehudit Meged, Avi Caciularu, Vered Shwartz, and Ido Dagan. 2020. Paraphrasing vs coreferring: Two sides of the same coin. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4897–4907, Online. Association for Computational Linguistics.

Jakob Nielsen. 1993. *Usability engineering*. Academic Press.

Jakob Nielsen. 1994. *Usability engineering*, page 165. Morgan Kaufmann.

NIST. 2005. DUC 2006: Task, Documents, and Measures. https://duc.nist.gov/duc2006/tasks.html. Accessed: 2020-05-19.

Brendan O'Connor, Michel Krieger, and David Ahn. 2010. TweetMotif: Exploratory Search and Topic Summarization for Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1).

Filip Radlinski and Nick Craswell. 2017. A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, page 117–126, New York, NY, USA. Association for Computing Machinery.

Siva Reddy, Danqi Chen, and Christopher D. Manning. 2019. {C}o{QA}: A Conversational Question Answering Challenge. *Transactions of the Association for Computational Linguistics*, 7:249–266.

Tuukka Ruotsalo, Giulio Jacucci, and Samuel Kaski. 2020. Interactive faceted query suggestion for exploratory search: Whole-session effectiveness and interaction engagement. *Journal of the Association for Information Science and Technology*, 71(7):742–756.

Ori Shapira, Ramakanth Pasunuru, Hadar Ronen, Mohit Bansal, Yael Amsterdamer, and Ido Dagan. 2021. Extending Multi-Document Summarization Evaluation to the Interactive Setting. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 657–677, Online. Association for Computational Linguistics.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised Open Information Extraction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 885–895, New Orleans, Louisiana. Association for Computational Linguistics.

Daniel Tunkelang. 2009. Faceted Search. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–80.

UIUX-Trend. 2021. Measuring and Interpreting System Usability Scale - UIUX Trend. https://uiuxtrend.com/measuring-system-usability-scale-sus/. Accessed: 2021-06-28.

MW Van Someren, YF Barnard, and JAC Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress*.

Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98.

Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. CorefQA: Coreference resolution as query-based span prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6953–6963, Online. Association for Computational Linguistics.

Yumo Xu and Mirella Lapata. 2020. Coarse-to-Fine Query Focused Multi-Document Summarization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3632–3645, Online. Association for Computational Linguistics.

Michihiro Yasunaga, Rui Zhang, Kshitij Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev. 2017. Graph-based Neural Multi-Document

291

Summarization. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 452–462, Vancouver, Canada. Association for Computational Linguistics.

# A    Implementation Details

## A.1    Interface

The frontend uses the reactjs library (`https://reactjs.org/`) and the bootstrap library (`https://getbootstrap.com/`).

## A.2    Backend

The backend service is written in python, using the tornado web server library (`https://www.tornadoweb.org`). The summarization model was downloaded from huggingface (`https://huggingface.co/facebook/bart-large-cnn`). The service is deployed on a Linux server with CPU only.

All coreference and proposition alignment models described in §3.1 are previously trained models. Links to these trained models are available in the project's GitHub.

For creating the CD coreference clusters for events with the fine-tuned CDLM model, we used two 32GB V100-SMX2 GPUs, for about 6 hours per topic. For creating the CD coreference clusters for entities, we used one 12GB TITAN Xp GPU, for about 5 minutes per topic. For creating the proposition alignment clusters we used four GeForce GTX 1080 Ti GPUs, for about 10 hours per topic.

**CD entity coreference merging step.**    As described in §3.1, our final CD entity coreference step merges WD and CD predictions. The Span-BERT WD model outputs clusters of coreferring mentions, while the CD entity model (Eirew et al., 2021) outputs a pairwise score for each pair of mentions. We therefore convert SpanBERT clusters to mention-pair scores, by scoring pairs that are clustered together as 1, and 0 otherwise. Then, following common practice (Kenyon-Dean et al., 2018; Barhom et al., 2019; Meged et al., 2020; Eirew et al., 2021), we apply agglomerative clustering over all mention-pairs (both WD and CD) and produce the final entity coreference clusters. Since WD coreference quality is superior to that of CD coreference, the high *WD* coreferring mention-pair scores of 1 causes the clustering algorithm to favor those pairs for overall coreference clusters.

**Proposition-level similarity threshold.**    The proposition alignment model computes a pairwise similarity between pairs of propositions, and we only consider pairs with a score above 0.5 (as a standard binary classification heuristic). We then create a similarity graph, where each proposition is a node, and paired propositions are linked with an edge. The final clusters are the connected components in the graph. For example, if for propositions $P_1$, $P_2$ and $P_3$, there exist pairs $(P_1, P_2)$ and $(P_1, P_3)$, then $P_1$, $P_2$ and $P_3$ will be clustered together.

**Facet-value label.**    As mentioned in §3.1, each facet-value is linked to a coreference cluster (a set of mentions) and has a label which is displayed to the user. For concepts and entities, this label is the text of the cluster's most frequent mention. For statements, there is no repetition of mention texts in the cluster. There, we use the text of the longest mention, under the assumption that it has more context for the user to understand the statement.

**Entities sub-facet categorization.**    After computing the *Entities* facet-values with entity coreference resolution, we categorize each facet-value to a specific entity type. For this, we first calculated the named entity class, with NER, for each mention in the facet-value cluster. All tokens of a mention were to be classified with the same NER class in order for the mention to be considered classified. Then, the class repeating the most times in a cluster was chosen as the class of the cluster. If all mentions of a cluster were not classified, we categorized the facet-value as *Miscellaneous*.

We mapped spaCy's NER classes to names that we found are more friendly to non NLP-practitioners (e.g., "GPE" is named "Location").

**Facet-value filters.**    After generating the potential facet-values (coreference clusters), we filter out:

- Clusters with more than 50 mentions, under the assumption that they are too noisy for the user.

- Singleton clusters, i.e. a cluster with one mention or one linked sentence (coreferring in the same sentence), under the assumption that they are uninformative.

- Clusters whose label is 2 characters or less (e.g., "'s", "AP").

- Clusters whose label has a verb part-of-speech tag.

**Summarization model input.** As described in §3.2, BART is used to summarize the set of input sentences relevant to the facet-value selections. Since BART has an input-length limit of 1024 tokens, ordering the sentences based on their sentence index raises the likelihood that summaries will be based on sentences from multiple documents. The documents were ordered by their alphanumeric file system order based on their document ID.

### A.3 Data

DUC 2006 MDS dataset is used for demonstrating the application, specifically with 6 topics: D0601, D0602, D0606, D0608, D0617, D0629.

## B Experiment Details

We carried out a usability study and a system comparison experiment (§4), as well as a summary quality evaluation (§3.2).

### B.1 Usability Study

For the usability study, six participants were gathered based on prior acquaintance. Each user had a 45 minutes Zoom session with an experienced experiment overseer. The participants first filled an experiment participation consent form. Before starting the actual experiment, the users were presented with another topic for experimenting with the system, followed by instructions of the experiment overseer, to reduce the learning curve of using the system for the first time.

Table 1 shows the two tasks that each user received. Participants conducted the experiments on the two topics in different orders.

**SUS questionnaire.** The SUS questionnaire (Brooke, 1995) was filled once by each user after both topics, with the following 10 questions:

1. I think that I would like to use this system frequently.

2. I found the system unnecessarily complex.

3. I thought the system was easy to use.

4. I think that I would need the support of a technical person to be able to use this system.

5. I found the various functions in this system were well integrated.

| Topic | Task |
|---|---|
| Native American Challenges (D0601) | As a junior reporter, you were assigned a task to read 25 documents about Native American Challenge and hand out a draft to a reporter who will write the actual report. For your draft, describe two / three challenges that Native American communities face. For each challenge, explain any possible causes, difficulties that arise, and things being done for or against. |
| EgyptAir Crash (D0617) | As a junior reporter, you were assigned a task to read 25 documents about the EgyptAir Crash and hand out a draft to a reporter who will write the actual report. Describe the crash and two theories around it. For each theory, describe who stands behind it, who opposes it and what are the claims supporting it. |

Table 1: The tasks that each user received in both usability study and comparison study. The tasks order was shuffled among the users.

6. I thought there was too much inconsistency in this system.

7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very cumbersome to use.

9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.

To calculate the SUS score, the following procedure is taken (Brooke, 1995): First sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1,3,5,7,and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU. SUS scores have a range of 0 to 100.

The final scores of the six participants are:

| User | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Score | 82.5 | 85 | 50 | 97.5 | 100 | 82.5 |

**Usefulness questionnaire.** After exploring each topic, the participants filled a questionnaire as follows:

- For the requirements of the given task, how useful was the Facets component between 1 (not useful at all) and 5 (very useful)?
- Overall, the summaries output by the system were: between 1 (I disagree) and 5 (I agree)
  - Coherent
  - Informative
  - Non-Redundant
  - Length was about right

The average and (StD) results on the 12 sessions (2 topic for 6 participants) are:

| System Aspect | Score |
|---|---|
| Facets quality | 4.3 (0.7) |
| Summ. coherence | 4.7 (0.5) |
| Summ. informativeness | 4.2 (1.1) |
| Summ. non-redundancy | 3.8 (1.0) |
| Summ. length is about right | 4.3 (0.9) |

**Comments raised by participants.** During the sessions, the experiment overseer collected comments and ideas for improvements raised by the participants. The consensus was that the summaries were very impressive, especially when realizing that they summarize many sentences from multiple documents, and that the *Concepts* and *Entities* facets were useful for navigating through the vast information. For improvement, suggestions included to reverse the order of the history list, to add a reset button of all filters and to move the facet-value frequency meter closer to the facet-value label. Some mentioned that the *Statements* facet was less useful, since it acts like a summary that is unnecessary with respect to the navigation process.

### B.2 System Comparison Experiment

For the comparative experiment, we gathered 4 graduate students from our NLP lab and gave them offline assignments which took about 45 minutes. At the beginning, each student was given a document of instructions describing iFACETSUM and DocFetcher, and were told to take a few minutes to play with each system on a third topic. Then each student was given a document with an assignment, with the same tasks as the usability study (Table 1). The participants were told to stop the exploration process once they felt satisfied with their outcome. There were 4 variants of the assignment document (one for each student), for all combinations of 2 systems and 2 topics, where a participant does not repeat the topic on both systems.

**Questionnaire.** After both topics, the users answered a comparative usability questionnaire, as mentioned in §4.2.

The average time for completing the assignment with DocFetcher was 16 minutes and the average for iFACETSUM was 15 minutes. We found that drafts written by participants using the two systems were comparable in informativeness, and importantly that the participants preferred iFACETSUM over the standard search approach (from questionnaire results and general comments).

### B.3 Summary Quality Assessment

To assess the quality of the summaries output by our system (using BART fine-tuned on a summarization task), we collected 5 output summaries from each of the 6 supported topics (30 summaries overall) by submitting random facet-value selections (one or more selections per summary). These selections yielded sentence sets (summarizer inputs) of varying sizes (2 to 47 sentences).

The summaries were rated for five standard summary readability criteria, as defined in (Dang, 2006), on a 1-to-5 Likert scale. Two of the authors rated all summaries, and then reconciliated on scores with a large (3 or more points) difference, in which case scores may have been slightly revised. In addition, we added a sixth aspect - "Factuality", which was assessed by binary scoring. For each of the 30 summaries, a single sampled summary sentence was scored 0 if any fact in it did not have evidence in the source text, and 1 otherwise (30 sentences tested). We found that many sentences were lightly paraphrased or were fusions of two sub-sentential extractions, yielding high factuality scores. Results appear in Table 2.

## C iFACETSUM Features and Sample Session

**Feature Explanations.** Some of the features of iFACETSUM, presented in §2, are further explained here:

| Summary Aspect | Score |
|---|---|
| Grammatically | 4.20 (0.94) |
| Non-redundancy | 4.58 (0.77) |
| Referential clarity | 4.10 (1.02) |
| Focus | 3.93 (1.12) |
| Structure & Coherence | 3.55 (1.14) |
| Factuality | 93.3% |

Table 2: Average and (StD.) scores of the summary evaluation ratings over 30 random summaries generated by the system, with a 1 (worst) to 5 (best) scale. For Factuality, the score is the percent of factual sentences (out of 30 sentences).

- Clicking "Original sentences" for a summary opens a pop-up window with the set of sentences used to generate the summary. The sentences are marked with mentions pertaining to the selected facet-values. They are grouped by their parent document and then listed in order of their position in their corresponding document. (Figure 3)

- Clicking a document title in the sentences pop-up opens another pop-up window with that document in full. The sentences from the parent pop-up are marked in red. (Figure 4)

- Clicking "History" opens a pop-up window with all the facet-value selections and resulting summaries from the current exploration session. (Figure 5)

- If a complete sentence from the summary has already been seen in a previous one, that sentence is tinted in purple. We found this useful given the summarization model's occasional extractiveness. (Figure 6)

**Facet-value examples.** We show in Table 3 some examples of facet-values and their mention clusters.

**Sample session.** In Table 4 we show the facet selections and resulting summaries from part of a session in the usability study.



Figure 3: The original sentences popup, which lists the sentences used to create the inquired summary.

| Facet | Facet-Value Examples |
|---|---|
| Concepts | "treaties" ("agreements", "deals", "treaty", "deal", "settlements", ...) |
| | "revenues" ("incomes", "profit") |
| Entities | "Makah" ("The Makah tribe", "The Olympic Peninsula tribe", ...) |
| | "the plane" ("the jet", "767", "EgyptAir Flight 990") |
| Statements | "Native Americans are leaving reservations and relocating in urbans areas" ("Indians now living in urban areas", "migration from the reseravtions continues", ...) |

Table 3: Examples of facet-values.

| Query | Summary |
|---|---|
| treaties (34 sentences) | Tribal leaders hope **settlement** will bring assets they need to upgrade reservation. Law requires tribes to reach **compact** with state in which reservation lies if it wants to open a casino. California does not allow gambling in the state, which has not allowed gambling in Nebraska. Florida, Kansas and Alabama have sued the U.S. Interior Department. |
| treaties, New York (5 sentences) | McCurn previously ruled that **New York** illegally acquired the Cayugas reservation land in 1795 and 1807. The state purchased it in violation of the 1790 Indian Trade and Intercourse Act, which required Congressional approval for all Indian land **transactions**. It was long-standing New York policy to assume authority over Indian land **deals** within its borders. |
| treaties, Florida (1 sentence) | In addition, **Florida**, Kansas and Alabama, trying to block the opening of Indian casinos within their borders, have sued the U.S. Interior Department with the aim of overturning new rules that allow the federal government to license tribal casinos in cases where states are reluctant to negotiate **compacts**. |

Table 4: A snippet of a sample iFACETSUM session. Words in bold are mentions of the selected facet-value(s) (e.g., "compact" is a mention of "treaties").

## Original Document



**Document: NYT19990214.0127**

**#0** American Indians are the victims of violent `crimes` at more than twice the national average and , unlike the situation among whites and blacks where the large majority of crime victims are of the same race as the perpetrators , 70 percent of those committing `crimes` against Indians are of a different race , according to the first comprehensive study of `crimes` involving Indians , which was released Sunday by the Justice Department .

**#1** The study found that the nation 's 2.3 million Indians were far more likely to be victims of violent crimes than members of any other racial group , and that the rate of violent crime experienced by Indian women is nearly 50 percent higher than that by black males .

**#2** A full 60 percent of the perpetrators of violent `crimes` against Indians were whites , according to the victims , while 29 percent of the offenders were

**#3** other Indians and 10 percent were described as black , the report said .

**#4** By contrast , other studies have shown that 69 percent of the perpetrators of violent `crimes` against whites are also white , and 81 percent of those committing violent crimes against blacks are themselves black .

**#5** `

**#6** ` This highlights what has been going on out there for 130 years , since the beginning of the reservation system , " said Sidney Harring , a professor of law at the City University of New York School of Law and an expert on Indian `crime` and criminal law .

**#7** Harring said much of the violence against Indians by other racial groups was attributable to racism and alcohol , `` with Indians being victimized by poor , drunken whites , people on the margins hurting each other . '

**#8** '

**#9** There are still high levels of prejudice against Indians in the West , where most Indians live , he said , and a culture that lives on the edges of Indian reservations `` that tolerates this violence , '' even

Figure 4: The original document popup, marking the sentences and mentions relevant to the *sentences popup* from which this document was requested. A document enables the user to get more context on the summary.

## History



**Query:** `treaties`

Tribal leaders hope `settlement` will bring assets they need to upgrade reservation. Law requires tribes to reach `compact` with state in which reservation lies if it wants to open a casino. California does not allow gambling in the state, which has refused to allow gambling anywhere. Florida, Kansas and Alabama are trying to block the opening of Indian casinos within their borders.

**Query:** `crime`

First comprehensive study of `crimes` involving Indians released by Justice Department. Study: 70 percent of those committing crimes against Indians are of a different race. Rate of violent crime experienced by Indian women is nearly 50 percent higher than that by black males. President Clinton proposes increasing the 1999 law-enforcement budget for the Bureau of Indian Affairs by $25 million.

**Query:** `poverty`

Clinton's visit is first to a reservation by a president since Franklin Roosevelt. 38 percent of Indian children aged 6 to 11 live in poverty, compared with 18 percent for U.S. children of all other races combined. Tribal leaders figured that slot machines would provide new revenue for the tribe's 1,200 members.

**Query:** `poverty` `Pine Ridge`

Housing Secretary Andrew Cuomo said `Pine Ridge` is a metaphor for the poverty tour. The reservation sits in Shannon County, the poorest census tract in the nation. 'This is generations of poverty on `the Pine Ridge reservation` , with very, very little progress,' Cuomo said. 'Very, verylittle progress,' he said.

Figure 5: The history popup, containing all interactions from the current session. For each interaction, the facet selections and corresponding summary are shown.



Figure 6: The sentence is tinted purple, indicating it was already extracted as part of a previous summary, relieving the user from reading it again.