IAHLT's Guidelines for Text Summarization

Inspired by Amir's GUM summarization rules found on <u>GUM summarization guidelines</u> <u>Corpling@GU Wiki</u>

Further reading: https://gucorpling.org/amir/pdf/ACL_2023_GUMSum.pdf

Objective: Generate high-quality summaries to train text summarization models.

Contents

| The Summarization | |
|--|---|
| Process | 2 |
| | |
| Best Practices for creating gold-standard | |
| summaries2 | |
| General | |
| Rules | |
| 2-3 | |
| Text Summarization and Subjectivity | 4 |
| Example 1: Summary length, and context of place and time | 5 |
| Example 2: Fidelity to the original text | 6 |
| Example 3: Inferior and inferred information | 7 |
| Genre Specific | |
| Guidelines | |
| 8-10 | |

The Summarization Process

- 1. **Familiarization**: Begin by thoroughly reading the article. Ensure you understand the primary themes and key details.
- 2. **Identification:** Pinpoint the article's central point or main message.
 - Tip: As you read, you might find it useful to mark or highlight important sentences or passages.
- 3. **Drafting**: Write a summary of the identified central point and crucial details. Ensure your summary:
 - Is concise and captures the essence of the article.
 - Adheres to the provided summarization rules.
- 4. **Proportionality**: While there isn't a strict limit on the summary length, it should ideally be proportional to the article's length. For instance, a short news piece might require only a few sentences, whereas a lengthy research article may need a more extended, yet concise, summary.

Best Practices for creating gold-standard summaries:

- Ensure objectivity by avoiding personal opinions, biases and implied information not included in the original article.
- Use clear and straightforward language.
- Maintain the original meaning and context of the article, and stick to the core message. Ensure the summary focuses on the central theme of the article without delving into tangential topics.
- Cross-check your summary against the original article to ensure no crucial details are missed.

General Rules

Create a Comprehensive Substitute for the Original Text:

- 1. The summary should serve as the best possible replacement for reading the entire text.
- 2. Presence of a Predicate: Every sentence in the abstract should have a predicate.
 - Good example (Hebrew): "הוא רופא" (He is a doctor) "רופא" acts as the predicate here.
 - Bad example: "בבית החולים" This phrase lacks a clear predicate and is incomplete.
- 3. Complete Sentences Only: Phrases that do not form complete sentences should be avoided in the abstract. Every sentence should be able to stand alone and convey a clear message or piece of information.
 - Good example: "המחקר החדש מדגים את ההשפעה של תזונה על בריאות הלב"."
 - Bad example: "ההשפעה של תזונה על בריאות." This is a fragment and lacks context.
- 4. Limit on Subordinate Clauses: Each sentence in the abstract can contain only one subordinate clause, with a maximum of two in special cases where clarity and coherence demand it. For example, given this article, here are two options:
 - Acceptable example: "ערכנו בדיקה לגבי המלונית "בייבי ליס", וראינו שמדובר במלון לכל דבר ".ועניין. שם זאת, תהינו האם החוויה שווה את המחיר
 - Less ideal example: "אנחנו ידענו שנלד ב-"ליס", ומבדיקה שערכנו לגבי המלונית בבית" וחדרים מפנקים החולים, בייבי ליס, ראינו שאמנם מדובר במלון לכל דבר ועניין, עם 3 ארוחות וחדרים מפנקים ".ומרווחים, אבל המחיר גרם לנו לתהות אם אכן החוויה שווה את זה
- 5. The goal is to allow rapid reading of the content while understanding its central message, ensuring fidelity to the original text.
- 6. The purpose of the summary is not to rephrase the content in different words or provide excessive information.

Note:

• For each rule, two summaries are provided as examples. The first, in green, adheres to our guidelines. The second, in red, contravenes them.

Text Summarization and Subjectivity

Defining concepts like "critical points" and "core to the main topic" can be challenging due to their inherently subjective nature. However, we can attempt to formalize them for clarity:

1. Critical Points:

• **Definition**: Essential pieces of information or facts without which the fundamental understanding or meaning of the content would be compromised.

• Criteria:

- **Relevance**: The point directly pertains to the main subject or thesis of the content.
- **Impact**: Removal of the point would significantly change the comprehension or takeaway of the content.
- **Uniqueness**: The point provides distinct information not reiterated elsewhere in the content.

2. Core to the Main Topic:

• **Definition**: Information that directly addresses the primary subject, theme, or purpose of the content.

• Criteria:

- **Subject Matter**: The information directly relates to the title or main theme introduced.
- **Central Idea**: The information is integral in developing, supporting, or concluding the primary argument or narrative.
- **Focus:** The information is frequently revisited or emphasized throughout the content.

In practice, discerning what is "critical" or "core" often requires a blend of analytical evaluation (based on the criteria above) and some level of subjective judgment, informed by the context, audience, and purpose of the content.

Example 1: Summary length, and context of place and time

רן ברץ, יועצו של רה"מ נתניהו, מסיים את תפקידו | ישראל היום Link:

| רן ברץ, יועץ בכיר לענייני תקשורת של ראש | רן ברץ, יועץ בכיר לענייני תקשורת של ראש הממשלה |
|--|--|
| הממשלה נתניהו, הודיע אתמול על סיום תפקידו. הוא | נתניהו, הודיע על סיום תפקידו. |
| מסביר שבחר לסיים את תפקידו מכיוון שהוא אינו ממצה | |
| את עצמו בתפקיד זה, ומכחיש את השמועות על יחסים | |
| עכורים בינו ובין נתניהו. | |
| | |

Issues with the Bad Summary:

- 1. **Inadequate Length**: A summary that's too short can often miss essential details. While brevity is important, it shouldn't come at the expense of leaving out crucial information.
- 2. **Omission of Time References**: Timeframes provide context. A good summary should include when an event or action took place, especially if it's significant to the understanding of the original content.

Recommendations:

- 1. **Incorporate Participant Names**: It's crucial to mention the names of the participants in the event. Specify who performed the action and who was affected by it. This provides clarity and helps in understanding the sequence and impact of events.
- 2. **Detail the Event:** Outline the primary components of the event, answering the key questions:
 - What happened?
 - **Where** did it take place?
 - When did it occur?

By paying attention to these aspects, you ensure that the summary is comprehensive and mirrors the essence and details of the original content.

Example 2: Fidelity to the original text

Link to the article: הולנד: עשרות אלפי חורפנים מומתים בגז

בעקבות הגילוי כי וירוס הקורונה עובר בין חורפנים ובני אדם, החליט בית המשפט בהולנד להמית את כל החורפנים שנדבקו בוירוס בחוות הגידול במדינה, למרות התנגדותם של פעילי בע"ח.

לאחר שנתגלה כי וירוס הקורונה בחורפנים יכול להדביק גם בני אדם, נדחתה עתירתם של פעילי בע"ח אשר דרשו להכניס את החיות שנדבקו בקורונה להסגר במקום להמיתם. בעקבות כך עובדי משרד הבריאות ההולנדי החלו בתהליך המתה בגז של עשרות אלפי חורפנים בחוות בדרום הולנד. התפרצות הקורונה בחוות הגידול לחורפנים העלתה מחדש את סוגיית הענף החקלאי של גידול חורפנים לתעשיית האופנה, ענף אשר ממשלת הולנד מתכוונת לסגור לחלוטין ב-2024.

Issues with the Bad Summary:

The summary includes many details that aren't central to the article. Additionally, it omits the mention of the key actor, in this case, the Dutch court.

Guidelines:

Ensure Fidelity to the Source:

- 1. **(a)** Do not introduce information that isn't present in the original text (that includes acronyms not specifically mentioned), even if it's implied or factually correct based on common knowledge.
- 2. **(b)** Refrain from drawing logical conclusions from content that the text alludes to but doesn't explicitly state.

Commentary

It's evident that the "bad example" overflows with details, not all of which are core to the main topic of the article. An effective summary should be succinct, covering the most critical points without delving too deeply into secondary or tertiary details, unless they're essential for clarity or understanding.

The guidelines provided highlight the importance of sticking to the original text's information and not making unwarranted assumptions or inferences. These are fundamental best practices to ensure that the gold data for model training is both accurate and effective.

Example 3: inferior and inferred information

Link to the article: יוסק'ה וחנה לא ידעו מה עלה בגורל בנם איתן - עד שהתבשרו כי נהרג בקרב גבורה בסיני • נחמה שרדה את השואה - ואיבדה את בנה מוטי • הסיפורים המרגשים

הורים מספרים על ילדיהם שנפלו במלחמת יום כיפור: יוסק'ה וחנה לוטנברג, הוריו של איתן לוטנברג ז"ל, ונחמה פוירשטיין, אמו של מוטי פוירשטיין ז"ל.

הוריהם של נופלים במלחמת יום הכיפורים מספרים על געגועיהם לילדיהם ועל החשש שזיכרון החללים דוהה והולך עם השנים, במיוחד של רווקים אמיצים וצעירים במותם כמו רב"ט איתן לוטנברג ז"ל, רב"ט מוטי פרוישטיין ז"ל ורב"ט מרדכי ינאי ז"ל, אשר קבורים בבית העלמין הצבאי בקריית שאול.

Issues with the Bad Summary:

- 1. The summary includes details not present in the original article (e.g., stating the fallen soldiers were brave when it isn't explicitly mentioned, even if it might be implied in the text).
- 2. It incorporates details that aren't central to the article's main topic.

Specific Emphases:

- 1. **Handling Typographical Errors**: If there are spelling errors in 'important' words, such as names of places/people (as in this example)
 - If it's clear it's a typographical error, correct it in the summary.
 - If the correct form is ambiguous (e.g., a family name could be "גולדמן" or "גודמן"), use a rule of thumb: Choose one form and use it consistently throughout the summary.
- 2. **Shell Nouns**: One can use shell nouns in article's reviewing something or in 'opinion' articles. Using "This article reviews.." or "The article's writer claims that..." will usually come handy in Israel Hayom articles.

Genre Specific Guidelines

- 1. **Bagatzim** Answer WH questions trying to follow the bagatz's natural order:
 - Include the parties of the case (Who)
 - Include the court and the case itself (Where, When and Why)
 - Include the parties debate (How)
 - Include the judge's name and verdict (what)

בקשה לרשות ערעור על פסק דינו של בית המשפט המחוזי בירושלים מיום 7.1.2021. נגד המבקש (עו"ד משה מאור) הוגש כתב קובלנה לבית הדין המשמעתי המחוזי של לשכת עורכי הדין מחוז תל-אביב והמרכז. ביום 15.8.2017 הורשע המבקש בעבירות שיוחסו לו בגין מעשיו. ביום 9.3.2020 גזר בית הדין המשמעתי המחוזי על המבקש עונש של הוצאה לצמיתות מחברותו בלשכת עורכי הדין, פיצוי למתלוננים בסך של 5,000 ש והוצאות למשיבה בסך של 5,000 ש"ח. המבקש ערער על הכרעת הדין ועל גזר הדין לבית הדין המשמעתי הארצי של לשכת עורכי לשכת עורכי הדין. לאחר ששמע את טיעוני הצדדים, ועל יסוד הסכמת המשיבה (ועדת האתיקה של לשכת עורכי הדין), קיבל בית הדין המשמעתי הארצי באופן חלקי את ערעור המבקש על הכרעת הדין בעניינו והורה על זיכויו מהעבירה של שליחת יד בכספי לקוח, אך לא מצא להתערב בהרשעתו ביתר העבירות. בית הדין הפחית את עונשו והעמידו על חמש שנות השעיה בפועל, חלף הוצאה לצמיתות מחברותו בלשכה, ופיצוי למתלוננים בסך של 5,000 ש. אופן ניהול ההליכים נגדו ונגד הקביעות של בית הדין המשמעתיים, כמו גם ביחס לשגיאות שנפלו, לטענתו, בפסק דינו של בית הדין המשמעתיים, כמו גם ביחס לשגיאות שנפלו, לטענתו, בפסק דינו של בית המשפט המחוזי. כבוד השופט מ' מזוז החליט כי דין הבקשה להידחות ועמה גם הבקשה לעיכוב ביצוע. כמו כן, כבוד השופט קבע כי המבקש יישא בהוצאות לטובת אוצר המדינה בסך 3,000 ש.

2. **Laundry lists** - Keep it short. For example, in an 'animals' article in which 20 animals are offered for adoption, one shouldn't include all of them nor their entire description in the article, but choose a few and be concise. Another example could be any article that reviews '10 Fun Hanukkah Activities', 'Best 12 vineyards this year', etc. In this case one should state 'This article reviews..." pick several items and describe them shortly.

פינת האימוץ השבועית: בראשון אוהבת חיות יום האימוץ מתקיים מדי שבת בין השעות 14:00-11:00. רבים מתקיים לאימוץ ביניהם פרינסס בת השנתיים וקוויל, כלב נאמן בן 4. בהרצליה ונתניה אוהבות חיות, יחכו לכם ליידי והאנה, פיטבוליות בנות 4 ו-2, לפרטים נוספים: 055-6675229. בצער בעלי חיים רמת גן והסביבה, חפץ חיים 4, תל אביב מחכים לכם טדי, כלב בן 10 חודשים שובב ואנרגטי ולילי דוברמנית בת שנה. לפרטים נוספים: 2313*.

פינת האימוץ השבועית: בראשון אוהבת חיות מחכה פרינסס - מקסימה ילידת מאי 2018, עדינה, תמימה ומלאת חן. כמהה לתשומת לב וליטופים, מחפשת את האחד שיאמר לה כמה היא טובה יפה וחברותית. קוויל - זכר יליד אוגוסט 2016, כלב חמוד וידידותי עם אנשים. מתוק עם פרווה נעימה שדורשת חיבוקים. קוויל בחור נאמן שישמח להגן ולאהוב את משפחתו. ג'נסיס - מתוקה ופוטוגנית, חברותית ומרשימה, אוהבת לשחק, אוהבת להתלוצץ, אוהבת להתפנק וכל זה לא ניתן לעשות בכלבייה, אלא רק בבית שידע להעריך ולהעניק לה והיא

תיתן בתמורה ים של אהבה. סבסטיאן - סבסטיאן זקוק לחשיפה נכונה וסבלנית, הוא חשדן ועוד לא החליט מה עמדתו לגבי אנשים, הוא זקוק להרעפת אהבה בכמויות רצוי מאנשים עם ניסיון בטיפול בכלבים שזקוקים ליד מכוונת. יום האימוץ מתקיים מדי שבת בין השעות מכוונת. יום האימוץ מתקיים לפרטים ולאימוץ: לימור: 14:00-11:00 . לפרטים 152-3423650

Another example:

אחרי ימי התפוח בדבש עם התקוות לשנה מתוקה, ואחרי ימי חשבון הנפש והתענית, מגיע החג הקליל יותר, זה שמקשט את נופי הארץ, עירוניים וכפריים, בשלל סוכות יפות וצבעוניות. השנה נמתחים ימי החג וחול המועד משני צידיו של סוף השבוע הבא, ובהתאם מוצע למטיילים שלל אטרקציות, הפעלות, סיורים ואירועים. הכתבה מרכזת את האירועים המתקיימים עוד השבוע, לדוגמה באתר החרמון תוקם הסוכה הגבוהה במדינה, ובחול המועד יתקיימו סיורים מודרכים המשלבים טבע, מורשת וגיאופוליטיקה. לצידם יעמדו לרשות המבקרים שאר פעילויות הקיץ של האתר, ובהן גלישה במדרון באבובי קיץ (מגיל 5), מזחלות אקסטרים, נסיעה מהירה במזחלת זוגית ועוד. הסיורים והכניסה לאתר ללא תשלום, המתקנים בתשלום שנע בין 40 ל-54 שקלים, אך ניתן לרכוש כרטיסים מוזלים באתר האינטרנט. בגן הלאומי ציפורי תתקיים בימים רביעי וחמישי, 23-22 בספטמבר, פעילות סוכות ואושפיזין בימי מסיק. - בסמטאות העיר העתיקה ציפורי יוצבו ארבע סוכות החג אתרוג, לולב, הדס וערבה - והמבקרים יוכלו לפגוש דמויות מתקופת המשנה: רבי יהודה הנשיא, רבי עקיבא, ריש לקיש ואלישע בן אבויה. בנוסף יתקיימו: מופע מוזיקלי לכל המשפחה, הדרכה בין שבילים ובוסתנים עתיקים המשקיפים לבקעת בית נטופה, ולכבוד ימי המסיק - גם מסיק זיתים. בצפון הנגב יתקיימו בימי חול המועד סוכות ארבעה סיורים מודרכים, בסבסוד משרד התיירות ורשות הטבע והגנים: מפגש שלושת הגבולות וחבל שלום, בשילוב קטיף עצמי וטעימות של פירות וירקות; טיול מיער בארי, כולל סיור חקלאי חווייתי בחממות האננס של משק 77; סיור רכוב בדרך הבשור התיכונה וביקור בחוות היענים; וסיור רכוב בדרך פרשי היאומַנְרִי, מגבעת הכלניות ועד הכניסה למושב ניר משה, שבסופו ביקור במבוך מרים, מבוד של צמחי תבלין

הכתבה מרכזת את האירועים המתקיימים עוד השבוע, לדוגמה באתר החרמון תוקם הסוכה הגבוהה במדינה, ובחול המועד יתקיימו סיורים מודרכים המשלבים טבע, מורשת וגיאופוליטיקה. לצידם יעמדו לרשות המבקרים שאר פעילויות הקיץ של האתר, ובהן גלישה במדרון באבובי קיץ (מגיל 5), מזחלות אקסטרים, נסיעה מהירה במזחלת זוגית ועוד. הסיורים והכניסה לאתר ללא תשלום, המתקנים בתשלום שנע בין 40 ל-54 שקלים, אך ניתן לרכוש כרטיסים מוזלים באתר האינטרנט. בגן הלאומי ציפורי תתקיים בימים רביעי וחמישי, 23-22 בספטמבר, פעילות סוכות ואושפיזין בימי מסיק. - בסמטאות העיר העתיקה ציפורי יוצבו ארבע סוכות החג אתרוג, לולב, הדס וערבה - והמבקרים יוכלו לפגוש דמויות מתקופת המשנה: רבי יהודה הנשיא, רבי עקיבא, ריש לקיש ואלישע בן אבויה. בנוסף יתקיימו: מופע מוזיקלי לכל המשפחה, הדרכה בין שבילים ובוסתנים עתיקים המשקיפים לבקעת בית נטופה, ולכבוד ימי המסיק - גם מסיק זיתים. בצפון הנגב יתקיימו בימי חול המועד סוכות ארבעה סיורים מודרכים, בסבסוד משרד התיירות ורשות הטבע והגנים: מפגש שלושת הגבולות וחבל שלום, בשילוב קטיף עצמי וטעימות של פירות וירקות; טיול מיער בארי, כולל סיור חקלאי חווייתי בחממות האננס של משק 77; סיור רכוב בדרך הבשור התיכונה וביקור בחוות היענים; וסיור רכוב בדרך פרשי היאומַנְרִי, מגבעת הכלניות ועד הכניסה למושב ניר משה, שבסופו ביקור במבוך מרים, מבוך של צמחי תבלין ובושם. לצד הסיורים, יתקיימו בצפון הנגב עוד עשרות פעילויות בימי החג: טיולי אופניים, סדנאות יצירה, ביקור באתר ובאתרי מורשת ועוד. פרטים והרשמה באתר הבשור.

ובושם. לצד הסיורים, יתקיימו בצפון הנגב עוד עשרות פעילויות בימי החג: טיולי אופניים, סדנאות יצירה, ביקור במוזיאונים ובאתרי מורשת ועוד. פרטים והרשמה באתר הבשור. מוקד כנפי קק"ל יקיים בפארק אגם ירוחם בימי החג פעילויות צפרות: סיור צפרות לעת שקיעה, שבמסגרתו יגלו המשתתפים, מצוידים במשקפות, את עולמן הקסום של הציפורים. הסיור מתאים לגילי 3 ומעלה ויתקיים ב-22 וב-23 בספטמבר בשעה 18:00 מחיר: 20 שקלים. ביקב טורא תוכלו לטעום ולשמוע על היין זוכה הפרסים של היקב. כאן יכלול הביקור טעימות יין והדרכה, לצד פלטת גבינות ולחמים. מחיר: 79 שקלים. בתיאום מראש: 1051-2427728

3. **Opinion articles** - One can begin the summary saying: 'The article's writer claims that...' unless the name of the author actually appears in the text, then we would like to use the author's name.

כותבת המאמר טוענת שפסיקת בג"ץ, לפיה אין לאפשר לימודים אקדמיים בהפרדה עבור החברה החרדית, למעט לימודי תואר ראשון, פוגעת בשילוב החרדים באקדמיה, מכיוון שכל עוד אין הפרדה נמנעת השכלה גבוהה מהציבור החרדי בכלל ומנשים חרדיות בפרט.

או

עומר ינקלביץ' טוענת שפסיקת בג"ץ, לפיה אין לאפשר לימודים אקדמיים בהפרדה עבור החברה החרדית, למעט לימודי תואר ראשון, פוגעת בשילוב החרדים באקדמיה, מכיוון שכל עוד אין הפרדה נמנעת השכלה גבוהה מהציבור החרדי בכלל ומנשים חרדיות בפרט.