

A Community-Driven Data-to-Text Platform for Football Match Summaries

Pedro Fernandes¹, Luís António Santos², Sérgio Nunes^{1*}

¹INESC TEC and Faculty of Engineering of the University of Porto
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

²Communication and Society Research Centre, University of Minho
Campus de Gualtar, 4710-057 Braga, Portugal

up201603846@fe.up.pt, lsantos@ics.uminho.pt, sergio.nunes@fe.up.pt

Abstract

Data-to-text systems offer a transformative approach to generating textual content in data-rich environments. This paper describes the architecture and deployment of Prosebot, a community-driven data-to-text platform tailored for generating textual summaries of football matches derived from match statistics. The system enhances the visibility of lower-tier matches, traditionally accessible only through data tables. Prosebot uses a template-based Natural Language Generation (NLG) module to generate initial drafts, which are subsequently refined by the reading community. Comprehensive evaluations, encompassing both human-mediated and automated assessments, were conducted to assess the system's efficacy. Analysis of the community-edited texts reveals that significant segments of the initial automated drafts are retained, suggesting their high quality and acceptance by the collaborators. Preliminary surveys conducted among platform users highlight a predominantly positive reception within the community.

Keywords: natural language generation, human-in-the-loop, data-to-text, journalism, NLP application

1. Introduction

*zerozero.pt*¹ is a prominent Portuguese sports portal, offering comprehensive access to match-related data and statistics, encompassing teams, players, goals, substitutions, fouls, and corners. Additionally, it provides news coverage for premier competitions, curated by a team of journalists. Notably, this coverage encompasses merely 1.4% of the matches cataloged in *zerozero.pt*'s database. To take advantage of the vast volume of data, *zerozero.pt* pioneered the development of Prosebot, an automated system designed to craft text-based match reports predicated on structured match data. Prosebot harnesses standard template-based Natural Language Generation (NLG) methodologies, drawing parallels with analogous tools employed in diverse sectors, including finance, meteorology, and sports journalism.

Developed collaboratively by *zerozero.pt*'s engineering team and journalists, Prosebot employs NLG techniques to produce initial drafts of match summaries, which are subsequently refined by the platform's community members prior to publication. This paper elaborates on the Prosebot platform, highlighting its role in engaging *zerozero.pt*'s vast community in the creation of match summaries using automatically generated texts as a starting point. Through this platform, *zerozero.pt* aims to expand

its coverage to encompass more matches and shed light on lesser-known teams and competitions. Additionally, the paper presents findings from a survey conducted among *zerozero.pt*'s journalists and community members, offering insights into their perspectives on the platform's effectiveness and the broader implications of automation in sports journalism.

The structure of this article is defined as follows: Section 2 identifies state-of-the-art and advancements related to NLG applications in journalism, collaborative work between humans and automated systems in text generation, and empirical studies exploring journalists' perceptions of automation tools. Section 3 presents the findings of a survey conducted with *zerozero.pt*'s journalist, aimed at discerning their stance on the platform and their perceptions regarding the use of automation in newsrooms. Section 4 describes the technical architecture of the Prosebot platform and its underlying data-to-text NLG module. Section 5 presents the results of a survey conducted with the community of readers to evaluate the impact of the platform. In the final section, we present the conclusions and future development prospects.

2. Related Work

2.1. NLG Applications in Journalism

Natural Language Generation (NLG) techniques have gained significant traction in journalism, ev-

*Corresponding author.

¹<http://zerozero.pt>

ident in both scholarly research and commercial applications. Notable academic contributions include PASS (van der Lee et al., 2017), Chen and Mooney (2008), Plachouras et al. (2016), Leppänen et al. (2017), and SumTime-Mousam (Reiter et al., 2005). On the commercial front, companies such as Arria², Automated Insights³ and Narrative Science⁴ have developed products like Arria Sports, WordSmith and Quill, respectively.

With regards to previous research, we highlight the PASS system (van der Lee et al., 2017). PASS employs a template-based approach to produce Dutch sports reports. A unique feature of PASS is its ability to generate dual reports for each game, tailoring the tone to the reader's team preference. This contrasts with Prosebot, which produces a singular, neutral-toned report. The data for PASS is sourced from Goal.com⁵, encompassing details such as participating teams, final scores, goal scorers, fouls, historical game results, and prior encounters between the teams. To evaluate PASS, a user-centric assessment was undertaken, wherein participants were presented with PASS-generated texts and queried about their clarity and fluency. The feedback indicated a favorable reception, with fluency and clarity scores surpassing the neutral benchmark of 4. An important distinction between PASS and Prosebot lies in their approach: while PASS autonomously crafts reports, Prosebot integrates feedback from *zerozero.pt*'s readers in the process.

There have also been reports of popular newspapers using machine-generated news articles, although information available about such systems is scarce. In 2007, the Los Angeles Times launched a blog about homicide reports that would eventually feature machine-written texts following a simple template (Young and Hermida, 2015). In 2014, the same newspaper used Quakebot to generate the first article for an earthquake (Oremus, 2014). In 2019, BBC managed to publish a news story for every constituency (a total of 650) in the UK on the general election night, using an NLG system (Molumby and Whitwell, 2019).

2.2. Post-Editing and “Machine-in-the-Loop”

Based on our research, the collaborative effort of humans and machines in crafting sports articles appears to be a novel concept. Nonetheless, we identified two analogous methodologies: post-editing and machine-in-the-loop story generation.

Reiter (2020) discusses the idea of editing an automatically generated text before it being released. He references the post-editing technique, predominantly employed in machine translation, and cites an experiment involving a weather forecast generator (Sripada et al., 2005). In this experiment, domain experts revised the generated forecasts before distribution to clients. Notably, Reiter observed that while post-edit analysis identified areas for system improvement, the post-editing itself didn't markedly elevate the text quality. He ultimately posits that the optimal collaboration between human and machine entails the machine producing a coherent and precise draft, which the human can then refine with nuanced modifications.

Additionally, there are various text-to-text NLG systems with the purpose of supporting individuals, in areas including machine translation (Vaswani et al., 2017; Sennrich et al., 2016; Bahdanau et al., 2015) and automatic grammar checking (Dale et al., 2012). A notable variant of these systems is the machine-in-the-loop story generation. This concept is a counterpart to the widely recognized “human-in-the-loop” approach. In the machine-in-the-loop paradigm, the machine augments human capabilities, whereas in the human-in-the-loop model, humans actively contribute to the training of machine-learning models by offering feedback, labeling examples, or proposing features (Clark et al., 2018). The general idea is the following: when creative writers encounter “writer's block”, they can leverage NLG systems to auto-generate sentences based on provided input. Examples of such systems include STORIUM (Akoury et al., 2020), Writing Buddy (Samuel et al., 2016) and Creative Help (Roemmele and Gordon, 2015).

STORIUM is a gamified collaborative story-writing platform, that transforms the writing process into an interactive game. Within each game session, one participant assumes the role of the narrator, while others embody individual characters. To incorporate the “machine-in-the-loop” story generation, the developers integrated a GPT-2 language model, which was trained on approximately five thousand STORIUM stories. For evaluation purposes, a web service was established to relay the model's outputs directly to the STORIUM platform. Users can generate a scene entry by pressing a button, which crafts the content based on the preceding text's context. Subsequently, users have the flexibility to modify the generated content by adding or removing elements as desired.

To quantify the extent of retained text, the authors introduced a metric named USER (User Story Edit Ratings). This metric, rooted in the Longest Common Subsequence (LCS) variant of ROUGE (Lin and Hovy, 2003), gauges the similarity between the original and edited texts. Upon applying this

²<https://www.arria.com>

³<https://automatedinsights.com>

⁴<https://narrativescience.com>

⁵<http://goal.com>

metric and gathering insights from user interviews, the authors deduced that users perceived the generated text as fluent, attributing this fluency to GPT-2's pre-training. Notably, while a significant portion of the generated content was often deleted by users, retention was higher when the content closely aligned with the overarching narrative. This observation was further corroborated by the Pearson's correlation between the USER scores and user evaluations. In a deeper analysis of the preserved content, the authors noted that around 30% of proper nouns remained intact, suggesting that users often retained elements like character names, which typically remain consistent throughout a narrative.

2.3. Journalists' Perceptions towards NLG

The perspective of news journalists on automated tools has been a topic of interest in prior research. [van der Kaa and Kraemer \(2014\)](#) investigated the trustworthiness and expertise of automatically generated articles in comparison to those written by humans. [van Dalen \(2012\)](#) analyzed a collection of blog posts and articles to discern journalists' views on the potential impact of automation on their profession. [Kunert \(2020\)](#) conducted interviews with German software providers and sports journalists to understand the adoption of automated tools and the anticipated shifts in sports reporting. Our survey, targeting *zerozero.pt*'s journalists, draws inspiration from these studies, and we contrast our findings with theirs in the subsequent section.

3. Survey of *zerozero.pt*'s Journalists

To inform the design of the platform, we solicited feedback from the journalists within *zerozero.pt*'s newsroom. This survey also aimed to gauge their perceptions regarding the influence of automated content generation tools. Conducted in January 2021, the survey garnered responses from fifteen (15) journalists. Ensuring confidentiality, the survey was designed to be entirely anonymous, comprising both 5-point Likert scale questions and open-ended queries.

In relation to the platform, the survey revealed that 86.7% of the journalists believed that summaries should undergo approval before publication. This contrasts with the views of the product managers, who, based on informal discussions, held the opposite stance. As further elaborated in [Section 4.5](#), *zerozero.pt* embodies a collaborative spirit, heavily relying on community contributions for various informational facets. In alignment with this spirit, the Prosebot platform aims to maintain this collaborative approach. However, in heeding the

journalists' concerns about maintaining journalistic integrity, the platform implements post-publication moderation. This ensures that content deemed inappropriate or not up to the website's standards will be promptly removed.

When inquired about the attribution for the summary, 73.3% of the journalists concurred that both the collaborators and Prosebot should be acknowledged as co-authors. Such an approach ensures transparency, allowing readers to be informed of all contributors to the text.

In the second part of the survey, we asked questions regarding the integration and implications of automated tools like Prosebot within the newsroom. When contrasting the capabilities of journalists with those of automated tools, a predominant sentiment emerged: while journalists viewed themselves as possessing greater flexibility and analytical prowess, they acknowledged the speed and expansive reach of automated content generation – specifically, in the context of football, this allows even minor competitions to benefit from textual coverage. A portion of the feedback expressed apprehensions about the potential reduction in the need for journalists in the future. However, a significant majority of respondents (66.6%) were confident that this eventuality is unlikely. These findings, detailed further in [Figure 1](#), largely resonate with the insights from [van Dalen \(2012\)](#). A notable deviation is the prevailing belief among *zerozero.pt*'s journalists that they won't be replaced. This sentiment might stem from their prior interactions with tools like Prosebot and the assurances from *zerozero.pt*'s project managers, who have consistently positioned Prosebot as a supplementary aid rather than a substitute.

In one of the open-ended questions, we asked how the introduction of these tools could change the role of a journalist, and most respondents highlighted the freedom to develop other types of content such as investigative work, although a few of them (3) do not foresee any significant change. In the last question, we inquired about the characteristics of the journalistic work that automatic tools would be unable to replicate. The respondents understand that these tools do not have access to the full context of a match, and do not have the sensitivity, creativity, and subjectivity to deal with the game's nuances. One particular answer even states that “*the sensitivity, in-depth analysis and interpretation of each journalist is irreplaceable*”.

4. Prosebot Platform

This section presents the architecture and functionalities of the Prosebot platform. We start by listing the data sources that feed into the system, followed by a description of the design and implementation of the templates module that enables the Natural

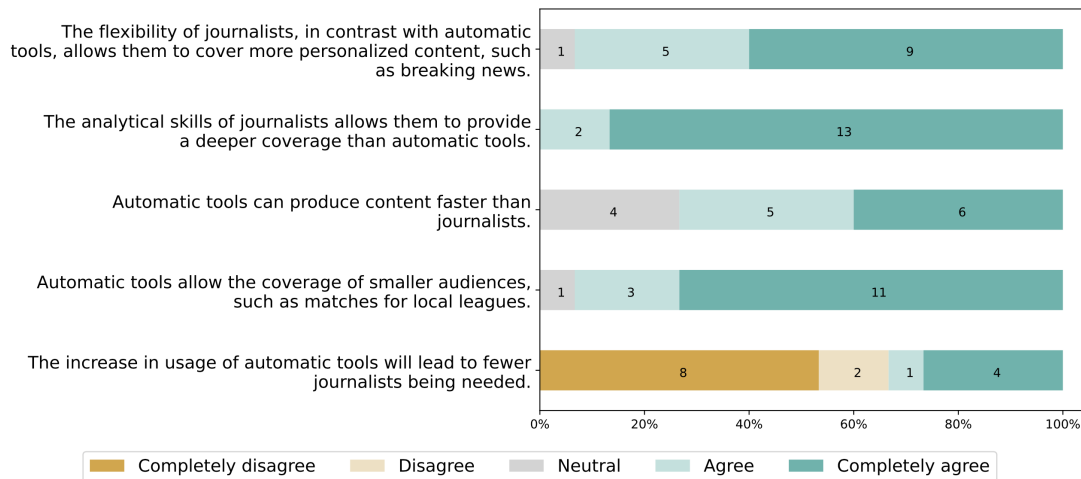


Figure 1: Overview of feedback from *zerozero.pt*'s journalists.

Language Generation (NLG) process. In the end, we describe the community-driven workflow that integrates readers into the content creation process.

4.1. Data Source

Contrary to other systems, such as those developed by Theune et al. (2001) and van der Lee et al. (2017), which sourced data through web scraping, Prosebot directly taps into the *zerozero.pt* database via a dedicated API. This API returns detailed match information, including participating teams, competition designation, final score, goal contributors, antecedent match outcomes, and player infractions. Furthermore, *zerozero.pt* also supplies data about the game's events and players' and teams' statistics. This allows for the generation of nuanced textual content, spotlighting, for instance, a player's scoring milestones or a consistent series of matches with goal achievements.

4.2. Template Design

In the current iteration of Prosebot, template have been developed through a collaborative effort between *zerozero.pt*'s engineering team and its journalists. These templates are archived in JSON files and are categorized for ease of access. Each template encompasses both content and an associated condition. Prosebot parses and evaluates this condition to decide if the content should be incorporated into the generated text. This approach has been particularly important in diversifying the template pool, thereby reducing textual monotony. Through a user-friendly graphical interface, journalists can autonomously modify and expand the existing templates. The most significant of these templates are detailed in Table 1.

4.3. Content Selection and Document Structure

The structure of the Prosebot-generated document closely mirrors that of news articles on *zerozero.pt*, with certain distinctions. The generated document is organized into seven primary sections. The Title provides a concise summary of the match outcome, be it a win or a draw. The Subtitle offers a brief characterization of the result, taking into account the goal difference, the context of goals scored, and the teams involved. The Small Text section details the match result, goal scorers, the number of red cards issued, and notable players. In the Introduction, the match result is highlighted, along with teams' performance trends leading up to the match, such as consecutive wins or losses. It also mentions the standout player and any significant trivia, which could include milestones like record-breaking shots or corners. The Events section chronicles key moments of the game, including goals, red cards, substitutions, and missed penalties. The Debriefing concludes with post-match team rankings, upcoming matches for each team, and potentially other trivia about each team. Lastly, the Trivia section, presented separately, lists intriguing facts about the match or the teams, encompassing sequences like winning streaks (e.g., "Team A broke a sequence of 5 winning games") and statistics, especially those that set new records for the competition.

Each section, with the exception of the final trivia section, is presented as a distinct paragraph. While all texts produced by Prosebot adhere to this format, the specific content can vary based on the match and the teams involved. For instance, in the context of a Champions League playoff match, the text would omit team rankings, as such information is not pertinent to that type of competition.

Section	Category	Variants
Title		All-purpose, 2+ goal difference, 4+ goal difference, home team won, away team won, game ended in draw, match ended with penalties
Subtitle		All-purpose, favourite team lost, decisive goal in the last minutes, no goals scored, many goals scored
Small text		Final score, starter and benched relevant players
Introduction		Final score, previous results, best player
Events	Goal	All-purpose, first goal, only goal, own goal, hat-trick, poker, second goal, last goal, goal drew the match, goal increased/decreased the goal difference
	Substitution	Gamechanger player is subbed in, relevant player is subbed in/out
	Missed penalty	Goalkeeper saved, penalty taker missed
	Red card	Direct red card, accumulation of yellow cards
Debriefing		Post-match classification, next games, match stats
Trivia	Stats	Best/worst result of the season for the team, best/worst overall result of the season
	Streaks	Increased or broke a sequence of matches

Table 1: Available templates.

4.4. NLG Algorithm

The NLG algorithm starts by parsing the template, grammar, and entity manager files based on the chosen language. At present, Prosebot is equipped to support three languages: Portuguese (both European and Brazilian variants), English, and Spanish. An example of a text generated in English can be found in Figure 2 in the appendix. The template files contain the textual content that will appear in the final result, according to the validity of their conditions. The grammar files cater to language-specific utilities, aiding in the representation of numbers in both ordinal and cardinal forms. They also facilitate the inclusion of articles based on the gender and plurality of the subject, a feature especially crucial for languages like Portuguese.

Entity managers primarily serve to retrieve names of entities, such as players or teams, based on stored information and previously used names. For instance, for a team like FC Porto, the system might be aware of its name, city, coach, and nickname. Consequently, throughout the generated text, the entity manager might alternate between references like “*Sérgio Conceição’s team*” and “*Porto’s team*” to denote FC Porto.

Upon loading the necessary files, the system aggregates data from the API request, categorizing it into appropriate classes. A pivotal class represents the match, serving as the primary access point for all data related to the match, teams, and players. With all data gathered, the system starts the text generation for each paragraph, as delineated in the preceding section. This generation largely adheres to a consistent strategy, with certain deviations tailored for the events paragraph.

The generation process is inherently recursive,

stemming from the multifaceted nature of the templates. A template can encompass not just textual content but also integrate functions, variables, and even other templates. The starting step involves selecting a root category. From this point, a subset of templates is used based on the fulfillment of their conditions. Given that each condition carries a specific weight, the filtering process also takes this weight into account, ensuring the selection of the most contextually apt templates.

To illustrate, consider two goal descriptions: “*After 20 minutes, Cristiano Ronaldo struck for Juventus*” and “*After 20 minutes, Cristiano Ronaldo scored Juventus’s only goal of the game*”. The former, being universally applicable to any goal scenario, lacks a specific condition. In contrast, the latter is contingent on the player’s team netting a solitary goal, thus bearing a higher weight. Consequently, in such a scenario, the latter template would be prioritized over the former.

Post the filtering phase, the system selects an appropriate template. Subsequently, a recursive function is invoked to parse the template’s content, interpolate variables, and if encountered, recursively process nested templates.

The generation process for the events paragraph diverges slightly from the standard approach. Instead of relying on a root template, the system’s objective is to chronicle the match’s events. To achieve this, it iteratively processes each match event. Leveraging the previously described generation process, it crafts text for each specific event and subsequently appends this to the overall narrative.

SC Braga defeated CD Tondela

Comfortable home win in goal rain

SC Braga defeated CD Tondela on Sunday, 4-2. Arsenalistas scored by João Novais, Lucas Piazón 2x and Ricardo Horta, while Tondela's team scored by João Jaquité and Souleymane Anne.

SC Braga triumphed over CD Tondela, 4-2, on Sunday, in the 20th round. In this competition, arsenalistas came from a win, and Tondela's team came from a win. [Lucas Piazón](#) was on fire. After 18 minutes, Lucas Piazón opened the scoring for SC Braga, with a right-foot shot inside the box, laid on by [Wenderson Galeno](#). After 40 minutes, [Ricardo Horta](#) fired home arsenalistas's second goal, laid on by Lucas Piazón. Shortly before the interval, [João Novais](#) struck for Braga's team, with a right-foot shot from outside the box, laid on by Lucas Piazón. With 50 minutes on the clock, Lucas Piazón struck for Carlos Carvalho's team, laid on by [Abel Ruiz](#). With 84 minutes on the clock, [Souleymane Anne](#) struck for CD Tondela, with a left-foot shot inside the box, laid on by [Salvador Agra](#). With 90 minutes already on the clock, [João Jaquité](#) netted the final goal of the game. After the result SC Braga are 3rd in the [table](#), 43 points, while CD Tondela occupy 12th place, 21 points. In their next fixture, arsenalistas [visit Nacional](#), while Tondela's team [will host Gil Vicente](#).

Figure 2: Sample of text generated by Prosebot in English.

4.5. A Community-based Platform

zerozero.pt has always been a proponent of community engagement, offering users the ability to register on the platform, participate in polls, comment on news and match pages, and even contribute valuable information about matches, teams, and players. Such contributions have been instrumental for *zerozero.pt* in providing coverage for a vast array of competitions, spanning from premier leagues to local tournaments. In line with this spirit, the Prosebot platform was conceptualized to empower *zerozero.pt*'s community members to partake in crafting match summaries, building upon an initial draft produced by Prosebot.

The workflow of the platform is straightforward and presented in [Figure 3](#). When a collaborator navigates to a match page, an option to craft a summary is presented. Upon selection, they are directed to a dedicated interface [1]. This interface features a text editor pre-populated with content generated by Prosebot's NLG module [2]. For reference, the original generated text is also displayed, allowing users to revert or cross-reference as needed. The page also showcases a trivia table related to the match and detailed event information. To guide collaborators, a brief overview of the initiative is provided, emphasizing certain guidelines. For instance, *zerozero.pt* advises against incorporating personal biases or subjective judgments into the summaries.

Upon submission of the summary [3], collaborators are presented with a brief survey, allowing them to share feedback on their user experience. Once processed, the summary is displayed at the top of the respective match page [4]. In the spirit of transparency, an alert is displayed alongside the summary, crediting both Prosebot and the contributing collaborator. The platform also facilitates a collaborative review process: other community members can comment on and rate the summary based on its quality and relevance [5]. Such feedback mechanisms are valuable for the moderation team, enabling them to identify and address potentially biased or inappropriate content.

5. Evaluation

In order to assess the efficacy and utility of the Prosebot platform, we undertook a multifaceted evaluation approach. Initially, we conducted an automated comparison between the initial drafts generated by Prosebot and the final versions that were published on *zerozero.pt*. This was to measure the degree of similarity and discern the nature of modifications made to the drafts. Complementing this, we sought direct feedback from users through a post-submission survey, aiming to capture their experiences and perceptions. Furthermore, we delved into an analysis of the data, examining the range of teams and competitions covered by the generated summaries and tracking the traffic these summaries attracted.

5.1. Post-submission Survey

To gain a deeper understanding of user experiences and perceptions, we administered a survey comprising both quantitative and qualitative components. The quantitative section consisted of seven (7) questions, each based on a 5-point Likert scale, where participants indicated their level of agreement with specific statements. The qualitative section, on the other hand, consisted of two open-ended questions, allowing respondents to provide more detailed feedback.

From the 46 responses we received, the general sentiment was overwhelmingly positive. As illustrated in [Figure 4](#), the mean score for all questions exceeded 4, indicating a high level of satisfaction among users. A notable observation was that a significant majority, approximately 81%, of the respondents had engaged with the platform more than once, suggesting a favorable inclination towards repeated use.

The open-ended questions provided insights into areas of potential improvement. One respondent highlighted a visual bug, while another pointed out design inconsistencies between the platform and the main website. A particularly interesting suggestion was the possibility of generating summaries even when complete match information wasn't avail-

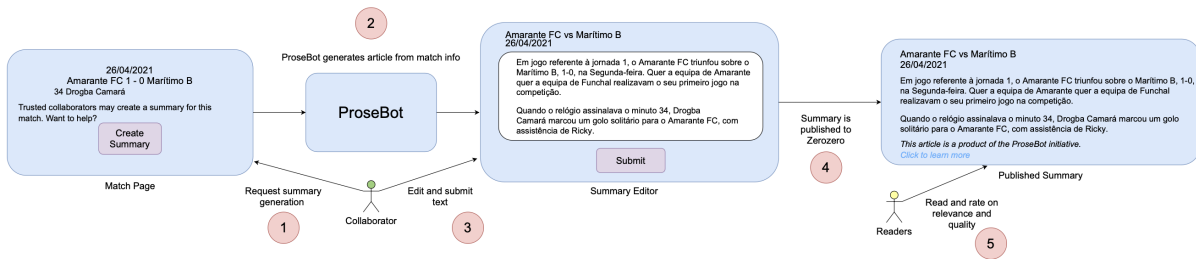


Figure 3: Overview of the Prosebot platform's workflow.

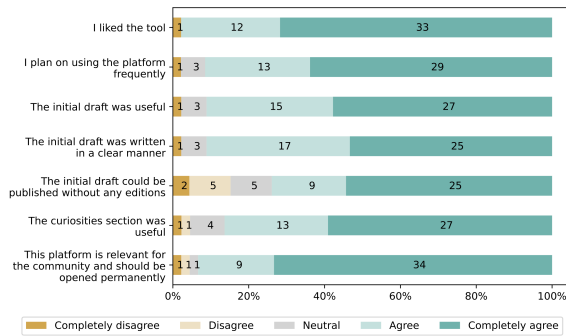


Figure 4: Survey to users results summary.

able. However, this poses challenges as incomplete match data would likely result in incomplete drafts, compromising the quality and comprehensiveness of the generated content.

5.2. Platform Usage and Engagement

The Prosebot platform was officially launched in early May 2021⁶, and the data presented here encapsulates the first month of its operation. A snapshot of key metrics can be found in Table 2.

In terms of user engagement, a majority of authors produced between 1 to 4 summaries. This suggests that while many users tried the platform, they might not have fully integrated it into their regular workflow. However, there's a smaller, highly engaged group of users who have been prolific in their contributions. Notably, the top five most active collaborators were responsible for a significant 40% of all summaries generated.

When examining the timeliness of the summaries, we observed that over half (53%) were published within two days of the match's conclusion, indicating a prompt response from collaborators. Interestingly, there's also a segment of users (18%) who chose to generate summaries for matches that occurred more than three months ago. This could be attributed to various reasons, such as nostalgia, historical interest, or the desire to fill gaps in coverage for older matches.

⁶<https://www.zerozero.pt/prosebot.php>

Indicator	May 2021	March 2024
Number of summaries	174	8,079
Number of authors	43	587
Number of teams covered	197	3,116
Number of competitions covered	56	295
Mean summaries per author	4.0	13.8
Mean summaries per team	1.8	5.1
Mean summaries per competition	3.1	27.4
Mean visits per summary	53.0	363.8

Table 2: Prosebot platform key statistics.

Revisiting these metrics three years post-launch, in March 2024 we can observe a clear consolidation of the Prosebot platform within the community (Table 2). During this period, more than 8 thousand match summaries have been generated by more than 580 users, covering over 3 thousand distinct teams. The top 5 users were responsible for 30% of all summaries generated and 14 users were responsible for 40% – confirming the initially observed pattern of the existence of a small number of very active users. Comparing the timeliness of summaries, 43% of summaries were published within two days of the match's conclusion. Notably, the platform has also been adapted for use in web portals in other countries, prominently in the United Kingdom and France, where it has been employed to generate match summaries in their respective languages.

5.3. Automatic Comparison

To assess the similarity between the initial drafts generated by Prosebot and the final published texts one month post-launch (174 texts), we employed two distinct metrics: the Dice coefficient (Dice, 1945) and a metric inspired by the approach used in STORIUM (Akoury et al., 2020).

The latter metric leverages Python's Sequence-Matcher⁷, a class tailored for sequence comparison. This algorithm identifies the longest contiguous matching subsequence, excluding elements like whitespace. This matching process is recursively

⁷<https://docs.python.org/3/library/difflib.html>

applied to the unmatched segments on either side of the identified subsequence. A specific method of this class, `get_op_codes`, details the transformations required to convert one document into another, using operations such as addition, deletion, replacement, or maintenance of text sequences. By utilizing this method, the metric captures the text segments retained by the user. The ratio of the cumulative length of these retained segments to the length of the original draft yields the percentage of text preserved from the initial draft.

However, this approach has a limitation. For instance, if a user merely rearranges the paragraphs without altering the content, the `get_op_codes` function would still interpret it as text addition and removal operations. To address this, we also incorporated the Dice coefficient. This metric computes the bigram similarity between two texts, irrespective of the sequence in which they appear, offering a more general measure of similarity.

Our analysis revealed that in a significant majority of instances, specifically in 82% of the summaries, over 90% of the content from the initial Prosebot-generated draft was retained in the final published version. When considering the Dice coefficient as a measure of similarity, the average score stood at 90%. Furthermore, 69% of the summaries boasted a Dice coefficient exceeding 90%, indicating a high degree of similarity between the drafts and the final texts.

An additional observation from our study pertained to the survey participants. We discerned a direct correlation between a user's satisfaction with the draft and the similarity of the draft to the final version. Specifically, the more a user felt that the draft was fit for direct publication without modifications, the closer the draft was to the final version in terms of content.

Finally, we have applied a manual and automatic word frequency analysis, to understand what kind of text the collaborator usually adds or removes. Collaborators frequently added depth to the final paragraph that discusses team classifications post-match. While Prosebot typically provides team positions in either ordinal or cardinal forms, collaborators often expand on this by noting if a team occupies the top or bottom position, if teams share the same points, or if there's been a change in the gap between teams. They also highlight significant achievements, such as a team clinching the league title.

In other instances, collaborators incorporate information that isn't readily available to Prosebot, such as specific goal opportunities or instances of rule violations. Furthermore, there's a tendency among some users to rephrase descriptions of goals, likely to offer a richer context or to present the information in a style they find more engaging or informative.

6. Conclusions

In this paper, we introduced the Prosebot platform, a pioneering effort in the Portuguese domain that integrates a community of readers into the post-editing process of automatically generated sports match summaries. By leveraging the power of data-to-text template-based systems, Prosebot provides initial drafts to *zerozero.pt*'s collaborators, who then refine and publish these summaries.

The platform's launch in May 2021 marked a significant milestone, and the subsequent feedback has been very positive. A survey conducted a month post-launch revealed a favorable reception among participants. Moreover, the analysis of user modifications to the generated drafts indicates a high retention of the original content, underscoring the system's capability to produce valuable drafts.

In the span of three years post-launch, Prosebot has been instrumental in generating millions of game summaries across various languages. This has notably amplified the coverage and visibility of matches at the grassroots level, which previously might have been overlooked.

This endeavor has not only showcased the potential of integrating NLG with community collaboration but also paved the way for future enhancements. By analyzing user modifications, we can continually refine and expand the NLG system's template repository, enhancing its diversity and adaptability. Furthermore, the modular nature of the NLG system offers the flexibility to extend the platform's support to other languages and even different sports, broadening its scope and impact.

Limitations

We have introduced a large-scale NLG platform designed to produce text summaries derived from football match statistics. The system is currently being used in a real-world setting with a user base of hundreds of thousands of users. Since the Prosebot system adopts a template-based strategy, one of the main limitations is the diversity of the generated text. Although a large collection of templates have already been developed, content diversity is impacted and can be improved. To tackle this problem, automatic template extraction is being explored to augment the number of available templates.

Ethics Statement

The data collected from platform users and journalists in the surveys was anonymized, ensuring their privacy protection. Participants provided informed consent for their responses to be used in research and analysis.

Automatically generated media, also known as *synthetic media*, is increasingly common. However, its societal impact is still under intense scrutiny and debate. In the design of the Prosebot system, actions were taken to control the impact of such systems. Most importantly, all game summaries generated have a clear disclaimer stating that the text was generated by Prosebot (seen on the right most box in [Figure 3](#)). Additionally, journalists were involved in the design of the system since early prototypes.

An ongoing discussion about the development of NLG systems is the impact on content production professionals, such as journalists. In the design of the Prosebot system we have found that initial fears of “machines replacing humans” have been replaced by a positive perspective of “machines helping writers”, as reflected on the survey the journalists reported in the paper.

7. Acknowledgements

We extend our sincere appreciation to ZOS for their ongoing and productive collaboration over the years, with special thanks to Marco Sousa and Pedro Dias.

This work is financed by national funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020 (DOI 10.54499/LA/P/0063/2020).

8. Bibliographical References

- Nader Akoury, Shufan Wang, Josh Whiting, Stephen Hood, Nanyun Peng, and Mohit Iyer. 2020. [STORIUM: A Dataset and Evaluation Platform for Machine-in-the-Loop Story Generation](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6470–6484, Online. Association for Computational Linguistics.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Nadine Braun, Martijn Goudbeek, and Emiel Kraemer. 2016. [The multilingual affective soccer corpus \(MASC\): Compiling a biased parallel corpus on soccer reportage in English, German and Dutch](#). In *Proceedings of the 9th International Natural Language Generation conference*, pages 74–78, Edinburgh, UK. Association for Computational Linguistics.
- David L. Chen and Raymond J. Mooney. 2008. [Learning to sportscast: A test of grounded language acquisition](#). In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, pages 128–135, New York, NY, USA. Association for Computing Machinery.
- Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. [Creative writing with a machine in the loop: Case studies on slogans and stories](#). In *23rd International Conference on Intelligent User Interfaces, IUI '18*, pages 329–340, New York, NY, USA. Association for Computing Machinery.
- Robert Dale, Ilya Anisimoff, and George Narroway. 2012. [HOO 2012: A report on the preposition and determiner error correction shared task](#). In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 54–62, Montréal, Canada. Association for Computational Linguistics.
- Lee R. Dice. 1945. [Measures of the amount of ecologic association between species](#). *Ecology*, 26(3):297–302.
- Jessica Kunert. 2020. Automation in sports reporting: Strategies of data providers, software providers, and media outlets. *Media and Communication*, 8(3):5–15.
- Leo Leppänen, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. [Data-driven news generation for automated journalism](#). In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 188–197, Santiago de Compostela, Spain. Association for Computational Linguistics.
- Chin-Yew Lin and Eduard Hovy. 2003. [Automatic evaluation of summaries using n-gram co-occurrence statistics](#). In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 150–157.
- Conor Molumby and Joe Whitwell. 2019. [General election 2019: Semi-automation makes it a night of 689 stories](#). *BBC News Labs*. Accessed 22 July 2023.
- Will Oremus. 2014. [The First News Report on the L.A. Earthquake Was Written by a Robot](#). *Slate*. Accessed 22 July 2023.
- Vassilis Plachouras, Charese Smiley, Hiroko Bretz, Ola Taylor, Jochen L. Leidner, Dezhao Song, and Frank Schilder. 2016. [Interacting with financial data using natural language](#). In *Proceedings of the 39th International ACM SIGIR Conference on*

- Research and Development in Information Retrieval*, SIGIR '16, pages 1121–1124, New York, NY, USA. Association for Computing Machinery.
- Ehud Reiter. 2020. [Human editing of NLG texts](#). *Ehud Reiter's Blog*. Accessed 22 July 2023.
- Ehud Reiter, Somayajulu Sripada, Jim Hunter, Jin Yu, and Ian Davy. 2005. [Choosing words in computer-generated weather forecasts](#). *Artificial Intelligence*, 167(1):137–169. Connecting Language to the World.
- Melissa Roemmele and Andrew S. Gordon. 2015. Creative help: A story writing assistant. In *Interactive Storytelling*, pages 81–92, Cham. Springer International Publishing.
- Ben Samuel, Michael Mateas, and Noah Wardrip-Fruin. 2016. The design of writing buddy: A mixed-initiative approach towards computational story collaboration. In *Interactive Storytelling*, pages 388–396, Cham. Springer International Publishing.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. [Neural machine translation of rare words with subword units](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.
- Somayajulu Sripada, Ehud Reiter, and Lezan Hawizy. 2005. [Evaluation of an NLG system using post-edit data: Lessons learnt](#). In *Proceedings of the Tenth European Workshop on Natural Language Generation (ENLG-05)*, Aberdeen, Scotland. Association for Computational Linguistics.
- M. Theune, E. Klabbers, J. R. de Pijper, E. Kraemer, and J. Odijk. 2001. [From data to speech: a general approach](#). *Natural Language Engineering*, 7(1):47–86.
- Arjen van Dalen. 2012. [The algorithms behind the headlines](#). *Journalism Practice*, 6(5-6):648–658.
- H.A.J. van der Kaa and E.J. Kraemer. 2014. Journalist versus news consumer: The perceived credibility of machine written news. In *Proceedings of the Computation+Journalism conference*. Computation + Journalism Symposium 2014 ; Conference date: 24-10-2014 Through 25-10-2014.
- Chris van der Lee, Emiel Kraemer, and Sander Wubben. 2017. [PASS: A Dutch data-to-text system for soccer, targeted towards specific audiences](#). In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 95–104, Santiago de Compostela, Spain. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Mary Lynn Young and Alfred Hermida. 2015. [From mr. and mrs. outlier to central tendencies](#). *Digital Journalism*, 3(3):381–397.