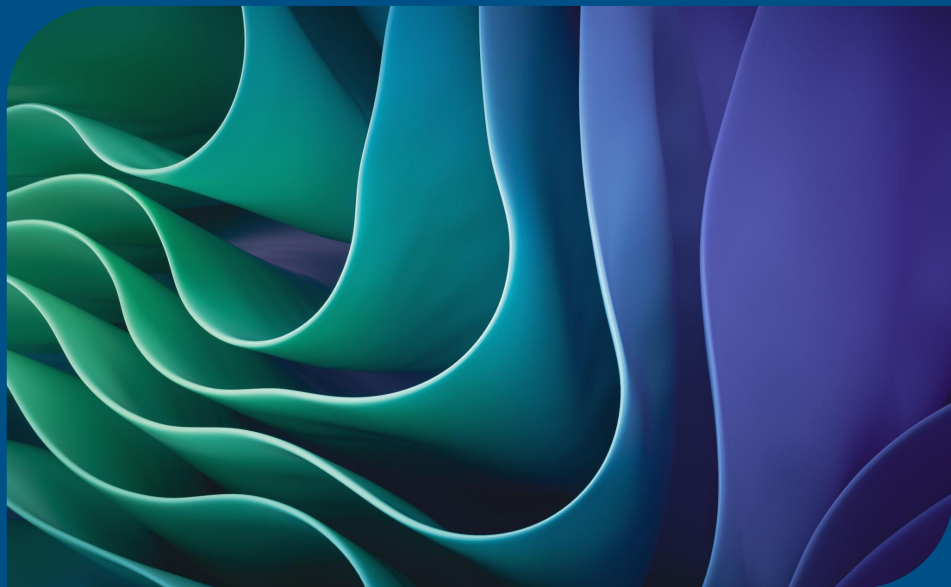# EPL Insight: Short-Form Match Summarization & Extraction

CSCI 4152/ 6509 – Natural Language Processing
Terry Quach | B00919525
Dalhousie

1. **Defining the Problem**

2. **Data Collection**

3. **Event Extraction**
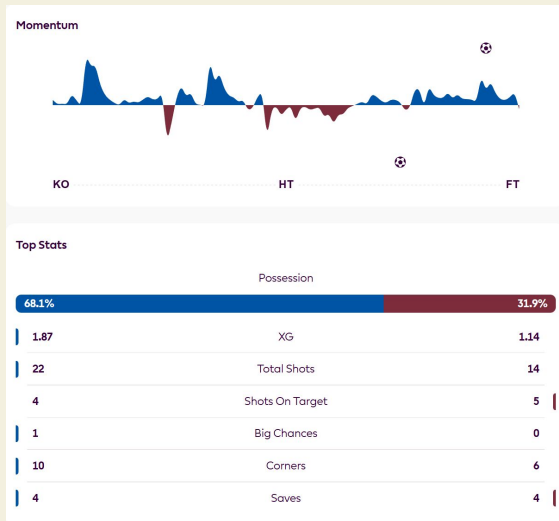
4. **Summarization**

5. **Drawing a Conclusion**

1. **Defining the Problem**

2. Data Collection

3. Event Extraction

4. Summarization

5. Drawing a Conclusion

Text-based "At-a-glance" soccer content

Summarization in sports is a very common topic, researched and attempted by many people and companies already (Zhang Y.).

But, like anything, just looking at the end result, doesn't tell the whole story.

And nowadays, the attention-span of people and their tolerance to read long reports or watch a whole 90 minute soccer game is dwindling (Mark).

1 - 1
HT 0 - 0

Brighton   **West Ham**

Rutter 90'+1'   ⚽   Bowen 73'

+

Matchweek 15
Sun 07 Dec   •   American Express Stadium   •   Simon Hooper
Brighton and Hove Albion vs West Ham United | Season 2025/2026 | Premier League

Momentum

KO   HT   FT

Top Stats

Possession
68.1%   31.9%

1.87	XG	1.14
22	Total Shots	14
4	Shots On Target	5
1	Big Chances	0
10	Corners	6
4	Saves	4

Report

Georginio Rutter's stoppage-time equaliser rescued a dramatic point for Brighton & Hove Albion in a 1-1 draw with West Ham United at Amex Stadium.

West Ham appeared set to climb out of the Premier League relegation zone after Jarrod Bowen slid in to put them ahead with 17 minutes remaining.

Brighton had generally struggled to generate a decent attacking threat for long periods of the contest, but piled on the pressure in the closing stages.

And in the first minute of stoppage time, it paid off, as Georginio drilled past Alphonse Areola to snatch a share of the spoils for the hosts.

West Ham stay 18th, two points from safety, while Fabian Hurzeler's side climb to seventh thanks to that late intervention.

How the match unfolded

Lucas Paqueta stung Bart Verbruggen's palms from distance during the first half, while, at the other end, El Hadji Malick Diouf was well-placed to block Georginio's goalbound volley.

Verbruggen then came to Brighton's rescue just before the hour mark, producing a superb double save to deny Bowen's deflected low drive and Crysencio Summerville's acrobatic follow-up.

However, the Seagulls goalkeeper was helpless as West Ham finally grabbed the breakthrough in the 73rd minute. Jan Paul van Hecke's loose header was pounced on by Callum Wilson, with the striker feeding a sliding Bowen, who got there ahead of Ferdi Kadioglou to steered the ball into the far corner.

Areola brilliantly tipped Mats Wieffer's deflected cross onto the bar soon after, as Brighton sought an immediate response.

The hosts' late persistence was eventually rewarded with an equaliser in the first minute of stoppage time, with Georginio drilling home from a tight angle after West Ham failed to deal with a corner.

There was still time for the visitors to regain their lead with the final action of the game, but Konstantinos Mavropanos nodded agonisingly wide from close range.

Brighton bounce back at the death

Following a generally toothless attacking display, Brighton somehow managed to drag themselves out of jail with a strong finish.

The Seagulls surrendered a 2-0 lead in their defeat to Aston Villa in midweek and appeared to still be reeling from that loss, as they struggled to truly threaten West Ham's goal.
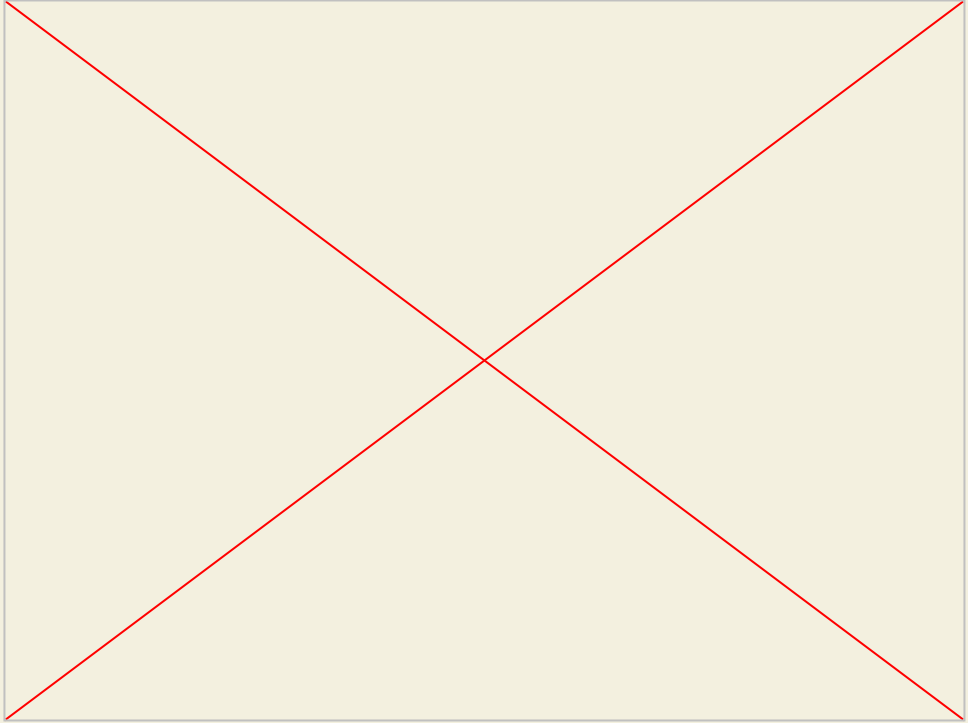
Brighton had Verbruggen to thank for keeping the scores level when he brilliantly denied both Bowen and Summerville in quick succession, getting back

1. **Defining the Problem**

2. **Data Collection**

3. **Event Extraction**

4. **Summarization**

5. **Drawing a Conclusion**

Scraping the website

I used Selenium & Python to scrape the English Premier League Website for the match score/details, stats and their report.

Once I cleaned up that Data, it was moved into a JSON file, ready for event extraction, summarization and manipulation

Scraping the website

I used Selenium & Python to scrape the English Premier League Website for the match score/details, stats and their report.

Once I cleaned up that Data, it was moved into a JSON file, ready for event extraction, summarization and manipulation

```
1   [
2       {
3           "home_team": "Brighton",
4           "away_team": "Leeds",
5           "final_score": {
6               "home": "3",
7               "away": "0"
8           },
9           "half_time_score": {
10              "home": "1",
11              "away": "0"
12          },
13          "scorers": [
14              {
15                  "team": "Brighton",
16                  "player": "Welbeck",
17                  "minute": "11'"
18              },
19              {
20                  "team": "Brighton",
21                  "player": "Gómez   , 70'",
22                  "minute": "64'"
23              }
24          ],
25          "cards": [...],
31          "stats": {...},
215         "report": "Danny Welbeck and Diego Gomez stole the show as Brighton
216     },
```

1. **Defining the Problem**

2. **Data Collection**

3. **Event Extraction**

4. **Summarization**

5. **Drawing a Conclusion**

All the other events

Named–entity filtering

Pattern rules

Regex

Context around verbs

Everton **3 - 0** Nott'm Forest
HT 2 - 0

Barry 45'+3' ⚽
Ndiaye 45'+3' (Assist)

Dewsbury-Hall 80'
O'Brien 80' (Assist)

Milenkovic 2' (OG) ⚽

Barry 57' 🟨   Savona 1'
Tarkowski 73'   Morato 45'+3'
Mykolenko 84'   Milenkovic 54'
O'Brien 90'+3'

One of those substitutes, **Ryan Yates**, then sustained what seemed to be a hamstring issue, and had to go off, further compounding Forest's frustration.

```
"events": [
    {
        "team": "Everton",
        "player": "Yates",
        "event": "Injury"
    },
```

9

1. **Defining the Problem**

2. **Data Collection**

3. **Event Extraction**

4. **Summarization**

5. **Drawing a Conclusion**

Summarization

Using BART Large CNN

Input Match Info

Split report into chunks to generate intermediate summaries

Humanize language

Then concatenate together into 3/4 sentence summaries

"raw_text": "Villa had shown some promise between the interval and Konsa's red card, though it was a largely underwhelming display from Unai Emery's team, who did not have a single shot in the first half.\nComing away with a share of the points at least showed Villa's resilience, while Newcastle may wonder what might have been had Isak been available.\nElanga could have got off to the ideal start in a Newcastle shirt when he raced through in the third minute, but Villa goalkeeper Marco Bizot - making his Premier League debut in the absence of the suspended Emiliano Martinez - made a fantastic save.\nGordon, filling in up top in place of Isak, headed over Harvey Barnes' cross soon after, before the England international saw a deflected attempt palmed wide by Bizot.\nIn the second half, Boubacar Kamara headed straight at Nick Pope after being picked out brilliantly by John McGinn, yet Villa soon found themselves up against it.\nElanga threaded a superb ball in behind for Gordon to run on to - Konsa dragged the forward back and referee Craig Pawson showed the first red card of the 2025/26 Premier League season.\nMatty Cash made a superb block from a fierce Barnes volley, while Tino Livramento cushioned a volleyed cross just out of Gordon's reach as Villa held firm.\nVilla won the corresponding fixture 4-1 last season, in what looked set to be a huge result in the race for Champions League qualification.\nBut it was not to be for Unai Emery's team, who missed out on a place at Europe's top table on the final day of last season, and they have been unable to strengthen their squad as much as Emery would have liked over the off-season.\nAnd they looked sluggish from the off, sloppily turning over possession and offering little coming forward in the first half.\nThey came to life after the break; Kamara might have done better with his header from McGinn's right-wing centre, while Ollie Watkins fired at Pope with his only real opening.\nJust as Villa were building up some pressure, though, it became a salvage job when Konsa got the wrong side of Gordon and reacted by hauling the attacker down.\nEmery will at least be pleased with the discipline and determination with which Villa defended from then on, while Bizot was in fine form when called upon, too. Substitute Donyell Malen offered some threat going the other way late on, but it would have very much been a smash-and-grab victory.\nHowever, Emery will want to see a much more well-rounded display from his team when they face Brentford next week.\nDespite the fact they have a UEFA Champions League campaign to look forward to, it has been something of a cruel summer for Newcastle.\nEddie Howe has had to answer plenty of questions over the future of star striker Isak, while he has seen reported targets slip through the club's grasp, too.\nYet Newcastle demonstrated they still have plenty of quality even without Isak - their fluid front three of Barnes, Gordon and Elanga were extremely effective. However, the lack of an out-and-out striker was also evident.\nOn his competitive debut for Newcastle, Elanga could have broken the deadlock. That chance was followed up by a penalty appeal, as Barnes went down, but there was nothing doing for the officials.\nGordon, too, squandered a couple of chances, and it is not difficult to imagine these kinds of opportunities having been put away by Isak, if he had featured.\nJust what happens now with Isak remains to be seen, but in the meantime, Howe will hope for more incomings. Jacob Ramsey is reportedly set to join from Villa in the coming days, though his arrival will not resolve the striker situation.\nNewcastle's first home game of the new campaign is against Liverpool on 25 August - an encounter that promises to be feisty.\nNewcastle have failed to win any of their last six Premier League games without Isak (D4 L2), failing to score in each of their last four, despite having 75 shots (9.4xG) across those six matches.\nVilla are unbeaten in their last 19 home matches in the Premier League (W11 D8), the Villans' longest ever run on home soil in the competition.\nNewcastle are winless in their last four away Premier League games (D2 L2), their longest streak in the competition since a run of seven between October 2023-January 2024. They've also failed to score in four of their last eight away league games, as many occasions as in their previous 25.\nVilla failed to have a single shot in the first half of a Premier League game for the first time since March 2022 v Arsenal, ending a 123-game run of managing a shot in the opening half in the league.\nAston Villa report | Newcastle report",

Aston Villa held on for a share of the points at St James' Park. Villa had shown some promise between the interval and Konsa's red card. Unai Emery's team did not have a single shot in the first half. Newcastle have failed to win any of their last six Premier League games without Isak (D4 L2) The Magpies have also failed to score Villa are unbeaten in their last 19 home matches in the Premier League

1. **Defining the Problem**

2. **Data Collection**

3. **Event Extraction**

4. **Summarization**

5. **Drawing a Conclusion**

Short-form Soccer Summary Success

More than just the scoreboard, but much more digestible than a whole report

Transition to other countries and languages (Wang, J)

Other sports (theanirudhvyas)

WEST HAM UNITED	1	-	2	Crystal Palace
West Ham				Crystal Palace
Jarrod Bowen 49'		⚽		Jean-Philippe Mateta 37'
				Tyrick Mitchell 68'

"West Ham were unable to deal with crosses in the first half. They almost fell further behind just over a minute after half-time. They were back in it through the indispensable Bowen, but it was ultimately not enough. The Eagles were boosted by the return of Adam Wharton, who had been missing for the past three weeks through injury. The hosts had been impressive in defence before conceding. "

Reference List

Premier League. (2025). Fixtures & matches. Retrieved from https://www.premierleague.com/en/matches

Slide 4:

Zhang, J., Yao, J.-g., & Wan, X. (2016). *Towards constructing sports news from live text commentary*. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 1361–1371). Association for Computational Linguistics. https://aclanthology.org/P16-1129.pdf

theanirudhvyas. (Year). *Sports-Commentary-Summarization* [Computer software]. GitHub. https://github.com/theanirudhvyas/Sports-Commentary-Summarization

American Psychological Association. (2023, February). *Why our attention spans are shrinking, with Gloria Mark, PhD* [Podcast episode]. In *Speaking of Psychology*. https://www.apa.org/news/podcasts/speaking-of-psychology/attention-spans

Slide 13

Wang, J., Li, Z., Zhang, T., Zheng, D., Qu, J., Liu, A., Zhao, L., & Chen, Z. (2021). *Knowledge Enhanced Sports Game Summarization*. arXiv. https://arxiv.org/pdf/2111.12535

Questions?