Chapter 5

Operationalised Stopping Strategies

In Section 3.2, we discussed a number of different stopping heuristics that have been defined in the literature. In this chapter, we take a number of these stopping heuristics forward to produce a series of different **stopping strategies,** providing an answer that addresses **HL-RQ2**.[1]

These stopping strategies are operationalised versions of their corresponding heuristics. This means that we can subsequently implement and evaluate their effectiveness. We consider twelve different stopping strategies across seven different categories, the categories being:

- **fixed depth**, which assumes a searcher examines to a fixed depth – and is also considered to be our baseline approach;

[1] Refer to Section 1.2 on page 10 for the definition of the research question.

- **frustration**, considering a searcher's *tolerance to non-relevance;*
- **satisfaction**, taking into consideration how satisfied a searcher feels with what they have found;
- **difference**, which operationalises how *different* new content appears to previously observed content;
- **IFT**, which considers a searcher's *instantaneous intake;*
- **time-based**, which utilise time as a measure for stopping; and
- **measure-based**, considering an established IR measure as a stopping strategy.

In the remainder of this chapter, we consider each of the seven categories enumerated above. For each category, we discuss the different operationalised stopping strategies that we use for the empirical work reported later in this thesis. Before this, we begin with a brief discussion about the different stopping decision points that were outlined in Section 4.2, and the notation used herein when describing the different stopping strategies.

Stopping Decision Points An open question that we have not yet addressed is that of what stopping decision points (of three presented in Section 4.2 on page 111) we will operationalise with the stopping strategies presented in this chapter.

For the purposes of this thesis, we consider the twelve operationalised stopping strategies purely in the context of **result summary stopping** – or considering the depth to which a searcher will examine a list of ranked results. The stopping strategies will be examined in tandem with SERP and session level stopping. These are left for implementation decisions as outlined in later chapters.

Selecting Stopping Heuristics Given all of the different stopping heuristics proposed in Section 3.2 beginning on page 78, a further open question about this work is: *how do you choose what heuristics to operationalise?* Stopping heuristics were selected that we believed

would offer good levels of performance for complex search tasks, where the onus was on the searcher to find and learn about a particular topic. Several of the reasoning-based stopping heuristics (such as the mental list heuristic, presented in Section 3.2.2.3) were not selected as operationalising them would have been too prohibitively complex.

A Note on Notation Each of the operationalised stopping strategies that are introduced in this chapter comes complete with at least one *stopping threshold* variable, allowing one to customise the point at which a searcher subscribing to a given stopping strategy should stop. As demonstrated in the **Presentational Conventions** front matter, the notation we use to illustrate a stopping strategy and its threshold(s) is NAME @THRESHOLD. As an example, SS1-FIX @3 denotes the fixed depth stopping strategy SS1-FIX, set to a threshold of 3. This stopping strategy is outlined below in Section 5.1.

5.1 Fixed Depth

The fixed depth stopping strategy is based upon an assumption held across many of the models and measures widely used throughout the IR community. The assumption is that a searcher will browse to a *fixed depth* before stopping when examining a list of ranked results. *P@k*, defined in Section 2.4.1.1, is a prime example of this, and has been used in many different studies examining the simulation of interaction. Given the wide use of this fixed depth approach in historical and contemporary IR and IIR research, we consider this stopping strategy as the baseline approach to which we will compare more advanced (and *adaptive*) stopping strategies.

- SS1-FIX **Fixed Depth** A searcher employing this stopping strategy will stop searching once they have observed x_1 result summaries (i.e. SS1-FIX @x1), regardless of the relevance of each judged result summary.

SS1-FIX is a naïve stopping strategy as it assumes that all documents up to rank x_1 are

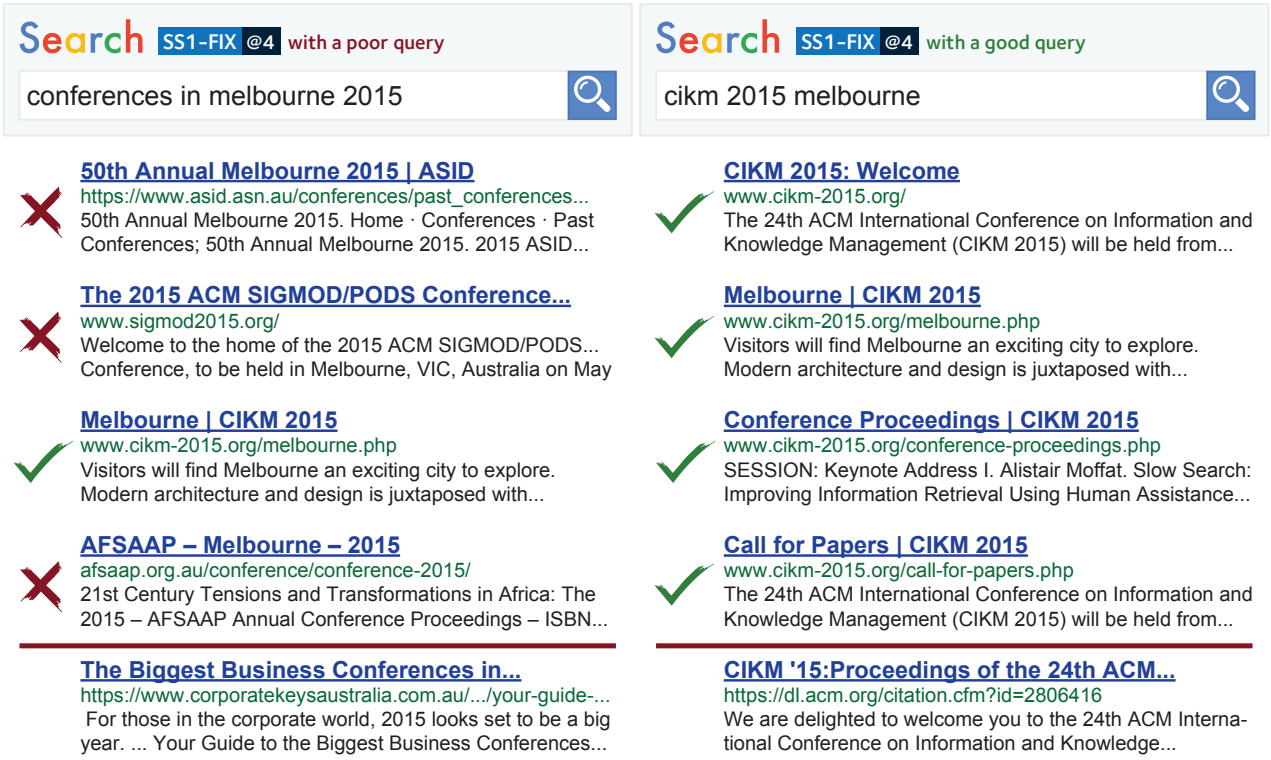

Figure 5.1 An example of the fixed depth stopping strategy, stylised in this thesis as **SS1-FIX**. Here, a searcher has an information need for the conference *CIKM 2015* in Melbourne, VIC, Australia. The left example shows the top five results for a poor performing query, with few unattractive results (denoted by ✗); conversely, the right shows results for a query performing well, with many attractive results (denoted by ✓). With **SS1-FIX @4**, the searcher will stop at a depth of *4*, regardless of the perceived relevance of the content provided.

considered attractive enough for a searcher to consider examining in closer detail. On average, this strategy does make sense. However, on a per-query basis, this strategy appears counter-intuitive and would be a waste of the searcher's time.

For example, Figure 5.1 demonstrates two SERPs side by side. Given a searcher's desire to find pages providing information to *CIKM 2015*[2], two queries are issued. The query on the left yields poorer results than the query on the right, denoted by the ✓ and ✗ that denote relevant and non-relevant results respectively. With **SS1-FIX @4**, four result

[2]CIKM 2015 was a conference held in Melbourne, VIC, Australia in October 2015. The paper that initially presented many of the different stopping strategies outlined in this chapter was presented by the author at that conference. Refer to Maxwell et al. (2015b).

summaries are always examined before stopping, regardless of the perceived quality of the results. Examining four documents for the query on the results list on the left is a waste of the searcher's time, with lots of non-relevant material. A searcher would be better *adapting* his or her behaviour depending upon the perceived quality of the ranked list.

5.2 Frustration and Satisfaction

We referred to SS1-FIX as a *fixed* stopping strategy, as it is not *adaptive*. The remaining stopping strategies presented in this chapter (with the exception of SS9-TIME) are considered to be adaptive as they permit a searcher to adapt their stopping depth depending upon the result summaries that they observe in a ranked list. In this section, we propose three adaptive stopping strategies that are based upon a searcher's *tolerance to non-relevance* and a simple *goal-based* approach.

5.2.1 Searcher Frustration

We first discuss how the frustration stopping heuristics are operationalised, as outlined in Section 3.2.1.1. Given a set of result summaries presented on a SERP, how many unattractive summaries would a searcher be prepared to examine before becoming frustrated with the SERP, and abandoning it? This stopping heuristic attempts to address this question. Indeed, as detailed in Section 3.2, a number of researchers have proposed stopping heuristics that consider unattractiveness.

The frustration heuristic intuitively makes sense for exhaustive searchers (Kraft and Lee, 1979). As an example, when tasked to find as many documents as possible related to different species of animals that are endangered, becoming disgusted with the presented results when a lack of unseen animal species are shown would be a suitable point at which to break and reformulate a new query, or abandon the search session altogether.

From the heuristics defined by Cooper (1973b) and Kraft and Lee (1979), we propose two variants of the frustration and disgust heuristics, SS2-NT and SS3-NC.

- SS2-NT Non-relevant, Total Under this stopping strategy, the searcher will stop once they have observed x_2 unattractive result summaries.
- SS3-NC Non-relevant, Contiguous Similar to the stopping strategy defined above, a searcher employing this stopping strategy will stop once they have observed x_3 unattractive result summaries *in a row (contiguously)*.

With these stopping strategies adaptable to the presented results, this inherently makes the strategies more realistic (Moffat et al., 2013). Figure 5.2 illustrates both strategies in action across the same query and associated results. On the left of the figure is an illustration of when a searcher employing SS2-NT would stop, and on the right, an example of SS3-NC. We use SS2-NT @3 and SS3-NC @3. Under SS2-NT, a searcher would stop at rank 5, while a searcher would stop at rank 7 when employing SS3-NC.

5.2.2 Goal/Satisfaction-Based

Analogous to frustration and disgust are the satisfaction, satiation and number-based stopping heuristics (Cooper, 1973b; Simon, 1955; Gibb, 1958). Rather than focus upon the frustration or disgust that a searcher might experience when confronted with unattractive result summaries, satisfaction-based stopping heuristics – explained in Section 3.2.1.1 – consider a searcher encountering a number of *attractive* result summaries before becoming sufficiently satisfied with what they have found before stopping.

- SS4-SAT Satiation A searcher using this stopping strategy will stop examining content after encountering x_4 attractive result summaries.

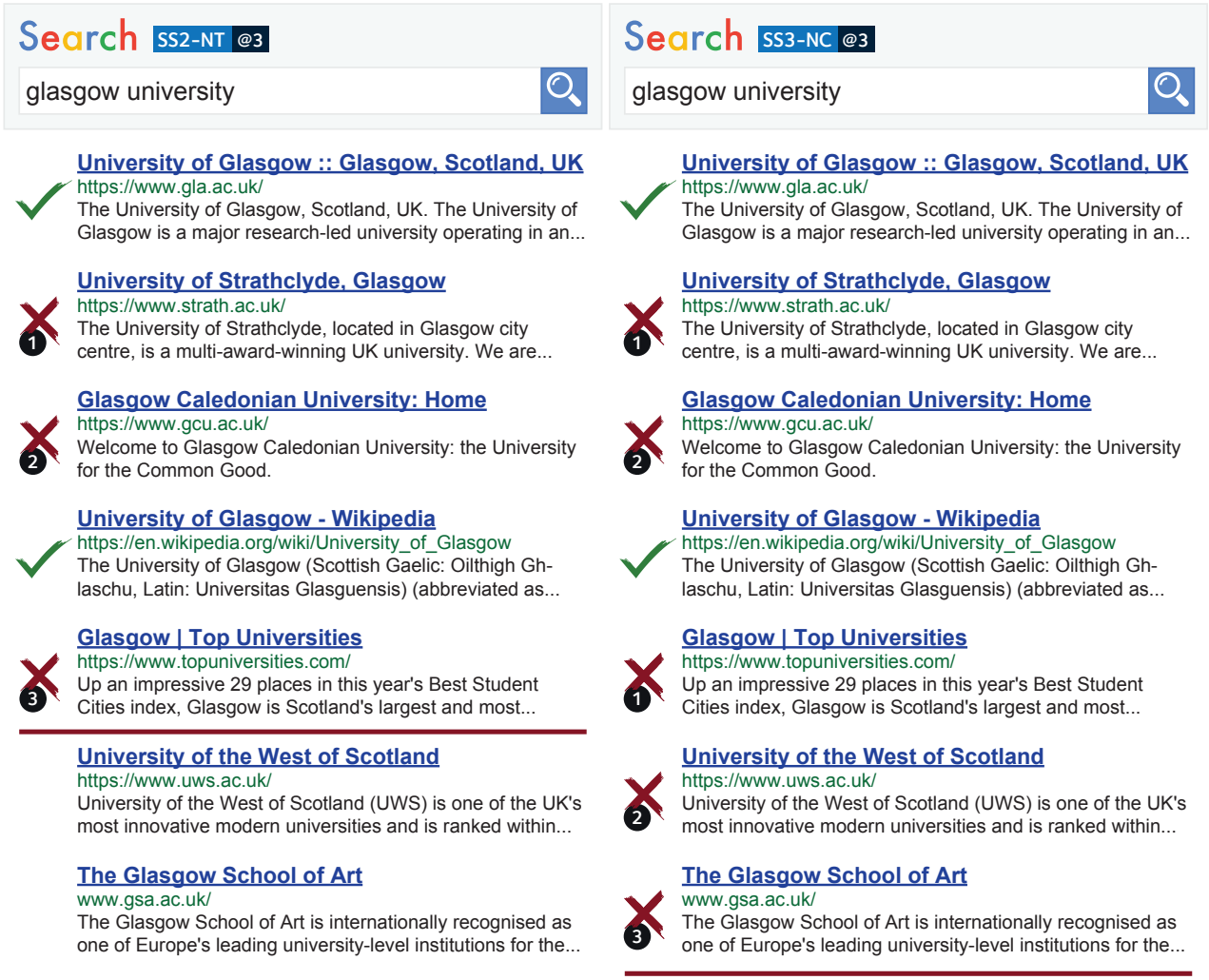

Figure 5.2 An example of the two frustration rules, **SS2-NT** (left) and **SS3-NC** (right), both three unhelpful (non-relevant) result summaries, under the same query and results. Given that **SS2-NT** considers the total number of result summaries judged to be unhelpful, a searcher employing this stopping strategy would stop at rank 5 in the example above. Considering a set of contiguous unhelpful summaries, a searcher using **SS3-NC** would stop at rank seven.

While we consider this stopping strategy in the context of result summary level stopping, such a stopping strategy may not be particularly useful when operationalised at this stopping decision point. Consider the scenario where a searcher issues a poor query, yielding next to no summaries deemed to be worthy of further examination. In this scenario, a searcher fully complying with **SS4-SAT** may struggle to find enough documents to reach their goal. This will mean that the searcher wastes time examining poor results. Such a stopping strategy may be better suited to an overall search goal – or at the session level

stopping decision point. As a means of potentially avoiding a searcher becoming *'trapped'* in an examination of a fruitless set of results, time limits could be imposed. We also consider an additional stopping strategy to alleviate this issue, as discussed below.

5.2.3 Combining Frustration and Satisfaction

The next stopping strategy proposed considered a combination of both the frustration/disgust and satisfaction/satiation stopping heuristics. This was named the *combination heuristic* by Kraft and Lee (1979). Employing this stopping strategy, a searcher would stop either when they became frustrated or were satisfied with the number of attractive summaries that they had observed – whichever of the two were met first. As such, we can convert this into a fifth stopping strategy, defined below.

- **SS5-COMB** **Combination — Frustration/Satiation** A searcher using this stopping strategy will employ both frustration (disgust) and satisfaction (satiation) stopping heuristics to determine when to stop, ceasing their examination of the SERPs contents for the first stopping heuristic whose criterion is met.

While **SS4-SAT** can be selected as the operationalised satisfaction/satiation component, one of either **SS2-NT** or **SS3-NC** can be selected for the frustration/disgust component of this fifth stopping strategy. We discuss this in our general methodology in Section 6.4.2.6 on page 173. Note that like **SS2-NT** and **SS3-NC**, we include items issued from previous queries of the same search session.

5.3 Difference Threshold

The next set of stopping strategies are based upon the difference threshold heuristic, as outlined in Section 3.2.1.2 on page 83. To operationalise this stopping heuristic, we considered

the difference between a given result summary's snippet text and the snippet texts of previously examined result summaries. Here, the idea was that as a searcher examined result summaries on a SERP, summaries may be encountered that are not *sufficiently different* from what had already been observed.[3] When encountering a result summary that is not sufficiently different, a searcher subscribing to the difference threshold stopping heuristic will then decide to stop examining results.

From this stopping heuristic, we devised two separate stopping strategies where the difference between snippet texts was computed in different ways. The first approach considered the *term overlap difference*.

- **SS6-DT** **Difference, Terms** This stopping strategy compares occurrences of terms in a given result summary's snippet text against all terms in previously examined result summary snippets. If $\frac{|s_{curr} \cup s_{prev}|}{|s_{curr}|} > x_6$, the new snippet text is then considered as too similar to previously examined result summaries. The searcher then stops examining result summaries on the present SERP.

Essentially, **SS6-DT** considers that if more terms overlap between old and new, the greater the chance that the new result summary would not contain any new information. In the definition above, s_{curr} denotes the terms of the currently examined result summary snippet, s_{prev} denotes terms from all previously observed result summary snippets[4], and x_6 is the threshold at which the searcher will stop.

The second difference-based stopping strategy utilised *Kullback–Leibler Divergence* (Kullback and Leibler, 1951) to determine how different a given result summary is from result summaries that have been previously examined.

[3] This means that searchers wouldn't be learning anything new (Nickles, 1995), and thus, under the eyes of such a strategy, would be wasting their time.

[4] All previously result summaries could be either session-based or query-based. This is an implementation decision, which we discuss in Section 6.4.2.6 on page 173.

- **SS7-DKL** **Difference, KL-Divergence** Here, KL-divergence is used as a means for comparing a given result summary (represented as a *bag of words*) against those previously observed. If the resulting value is less than threshold x_7, the present result summary is considered to be too similar, and the searcher stops. The searcher then abandons the present SERP.

Details related to the implementation of the difference heuristic stopping strategies, as well as the parameter threshold settings used, can be found in Section 6.4.2.6.

5.4 Instantaneous Intake

In Section 3.3.1.2, we discussed several stopping heuristics that were derived from OFT and IFT. The IFT-based heuristic considers a searcher's *optimal stopping point* at which a forager[5] should stop, as suggested by the underlying models of IFT. This is calculated by observing a searcher's *average rate of gain*. If the value of knowledge gained drops below this threshold, the searcher should stop, as graphically illustrated in Figure 3.8 on page 94.

We now propose an eighth stopping strategy, this time based upon the notion of the average rate of gain accrued by a searcher.

- **SS8-IFT** **Optimal Stopping** With this stopping strategy, a searcher is assumed to have some idea of the average rate of gain (denoted as x_8). If the rate of gain from the observed documents thus far does not exceed x_8, the searcher then stops and proceeds to undertake the next action as dictated by the CSM.

Computing the average rate of gain is a non-trivial problem. We leave specific implementation details of how this was achieved – along with other implementation details of the stopping strategy – to our methodology, reported in Section 6.4.2.6 on 173.

[5] As we discussed in Section 3.3.1, a *forager* can be considered analogous to a searcher seeking information.

5.5 Time-Based

In addition to the optimal stopping point approach discussed above, Section 3.3.1.2 also outlined a number of different OFT-inspired stopping heuristics that primarily used time as a measure of determining when to stop. From these approaches, we create two further time-based stopping strategies.

- **SS9-TIME** **Time-based** Based upon the *time heuristic* (Charles-Dominique and Martin, 1972; Krebs, 1973), a simulated searcher using this stopping strategy will abandon a SERP after x_9 seconds have elapsed since they entered it.
- **SS10-RELTIME** **Time, Give-Up** Using the *give-up heuristic* as defined by Krebs et al. (1974), a searcher will abandon a presented SERP x_{10} seconds after the last document that was found and considered relevant/useful (saved) to the given information need.

Given these stopping strategy definitions, **SS9-TIME** performs akin to **SS1-FIX**, in the sense it offers a fixed interaction time on each SERP, and is agnostic of the quality of the presented ranked list. Conversely, **SS10-RELTIME** offers a more adaptive solution similar to **SS2-NT** and **SS3-NC**, basing the time at which the searcher stops x_{10} seconds after a relevant document was last saved.

For this thesis, we also consider the *combination heuristic* proposed by McNair (1982). The stopping strategy that we propose based upon this heuristic assumes that a searcher has been able to acquire an idea of how potentially relevant summaries are *distributed* across the results presented within the SERP.

- **SS11-COMB** **Combination — Time and satiation** Encountering a SERP expected to yield a high volume of relevant content early on (high scent), a searcher will employ the satisfaction/satiation stopping heuristic. However, if the SERP is judged to yield

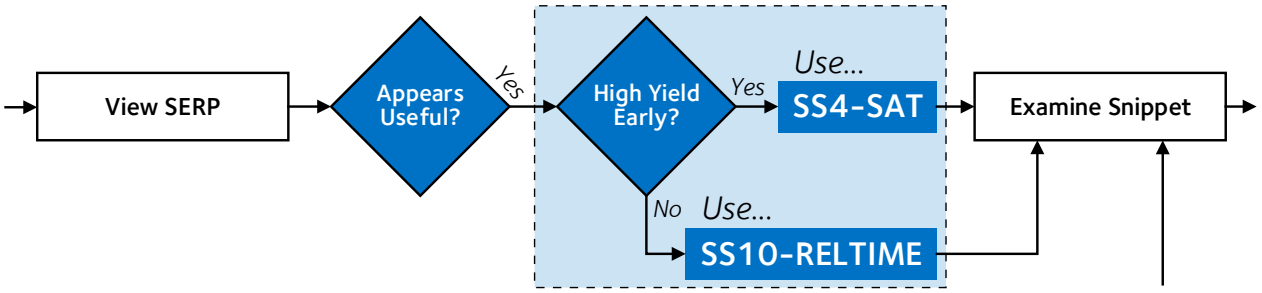

Figure 5.3 An excerpt of the CSM with the additional decision point that **SS11-COMB** incorporates within the searcher model. After deciding that individual result summaries within a SERP are worth examining in more detail, a searcher will then also have to decide whether the presented SERP will yield a high number of fruitful results *early* in the rankings, or trickle relevant material over greater depths (or not at all). The additional decision point and selected stopping strategies are highlighted within a blue box.

relevant items over greater depths or is judged to be of poor quality (low scent), the give-up time-based heuristic is used instead.

From our definitions above, **SS4-SAT** is used for the satisfaction/satiation component, and **SS10-RELTIME** is used for the give-up time heuristic component. The combination stopping strategy attempts to ensure that a searcher does not waste time on a SERP that appears to offer a low yield, but conversely capitalises upon patches that present a high yield. Of course, determining the perceived yield is a question of implementation; refer to Section 6.4.2.6 for more information on how we implemented this particular stopping strategy. Essentially, this combination stopping strategy incorporates an additional decision point within the searcher model, where one must determine if the presented SERP is high yield early on or not. This is illustrated as an excerpt of a flowchart in Figure 5.3.

5.6 Measure-Based

The final proposed stopping strategy is based upon an established IR measure. *Rank Biased Precision (RBP)* – as discussed in Section 2.4.1.5 – is utilised as the basis of our final stopping

strategy. Under RBP, the decision to continue to the next result in a ranked list is based upon a patience parameter or the probability of continuing. Essentially, RBP states that the probability of continuing decreases as a searcher progresses further down a ranked list.

- **SS12-RBP** **Rank-Biased Precision** With this stopping strategy, a searcher will stop examining a SERP when the likelihood of continuing falls below the RBP probability computed at that rank, given a patience parameter x_{12}.

By including such a measure, we provide a platform for which contemporary IR measures can be compared against the performance of other stopping heuristics defined in the literature. Implementation details, such as how we implemented the probabilistic component, can be found in Section 6.4.2.6.

5.7 Chapter Summary

This chapter has outlined 12 different stopping strategies, all of which are based upon prior stopping heuristics and an established IR measure. As such, this chapter provides a possible answer to **HL-RQ2**. In subsequent chapters of this thesis, we take these stopping strategies forward, discuss the specifics of how they were implemented in Section 6.4.2.6 (page 173), and how they were employed in our empirical experimentation.