Table 1. Key symbols and notations used in our algorithm description.

Symbol	Meaning
wp, wp_k	An arbitrary webpage without or with a reference index number
Ω	A certain user information need
Φ	A predictive model for webpage utility scores
$\Phi(wp, \Omega)$	The predicted utility score for wp according to Ω using text features extracted from the full text of wp after wp is downloaded
$T_1(wp)$	Content words or word phrases in the main body text of wp
$T_2(wp)$	Content words or word phrases in the heading and subtitle text of wp
$T_3(wp)$	Content words or word phrases in the anchor text embedded in wp
$kw_{i,j}(wp)$	Key words or phrases extracted from the text set $T_i(wp)$ $(i = 1, 2, 3)$ where j is an entry index number
n_i	The number of distinct keywords extracted by the RAKE algorithm from the text set $T_i(wp)$
$kw_{i,j}^k$	A short notation of $kw_{i,j}(wp_k)$
$\mathbf{wp}$	A selected collection of webpages where every webpage's relevance to a given Ω is manually rated
$\mathbf{kw}$	All detected keywords from webpages in $\mathbf{wp}$
$\hat{\Phi}(wp_\kappa, \Omega)$	A human rated utility score for the webpage wp_κ with respect to a given information need Ω
$\overline{\mathbf{kw}}$	A filtered version of the keyword set $\mathbf{kw}$ where infrequent keywords are removed
τ	A lower threshold for human labeled webpage utility scores
$\psi\big(kw_{i,j}^k, \Omega, \tau, \mathbf{wp}\big)$	The keyword $kw_{i,j}^k$'s odd ratio with respect to the labeled training set $\mathbf{wp}$ and the information need Ω
$\mathcal{C}_1$	A straightforward construction of a web crawler
$\mathcal{C}_2$	Our proposed adaptive user-oriented web crawler
$\{Q_1, Q_2, \cdots, Q_n\}$	A series of seed queries to obtain initial search result webpages for a web crawling session
u_i	A search result URL link
$\mathbf{u}$	A list of search result URL links
$\Delta(\mathbf{u})$	A prioritized crawling list for the URL list $\mathbf{u}$
$\delta()$	A webpage ranking function
$u_{\delta(i)}$	The URL of the i-th webpage for the crawler to visit in a given crawling session
$\mathbf{F}(u_i)$	Text features extracted from the webpage pointed to by the link u_i.
$\Upsilon(wp(u_i), \Omega, t_j)$	The predicted utility score for the webpage $wp(u_i)$ according to the information need Ω using all available information at the time moment t_j
t	A given time moment
$T(u_i)$	The time cost to access the webpage pointed to by the URL link u_i
$[-\psi_k(wp(u_i), t_j), \psi_k(wp(u_i), t_j)]$	$\Upsilon(wp(u_i), \Omega, t_j)$'s relative prediction error interval
$\eta_k\big(wp(u_i), t_j\big)$	Estimation confidence for the relative error interval of $[-\psi_k(wp(u_i), t_j), \psi_k(wp(u_i), t_j)]$
t_x	A certain amount of time permissible for a crawler
$\mathcal{V}(t_0, t_x)$	An optimal URL visitation trajectory for the crawling session starting at the time t_0 and ends at t_x
$\{u_{\delta(1)}, u_{\delta(2)}, \cdots, u_{\delta(n_x)}\}$	URLs of webpages crawled following a certain trajectory $\mathcal{V}(t_x)$ and within the time duration $[t_0, t_x]$
$\mathcal{U}(t_i)$	The snapshot of the pool of candidate URLs awaiting to be crawled at the time moment of t_i
t_0	The starting time for a crawling session
$\mathcal{U}(t_0)$	The pool of seed webpage URLs for launching the crawler
$\mathcal{G}\big(t_0, t_x, \mathcal{V}(t_0, t_x)\big)$	A theoretical objective function for optimal crawling trajectory planning
$\hat{\mathcal{G}}\big(t_0, t_x, \mathcal{V}(t_0, t_x)\big)$	A practical objective function for optimal crawling trajectory planning that can be computed on-the-fly
$p_i(t_z)$	The probability that from a candidate URL pool $\mathcal{U}(t_z)$ a link u_i is randomly selected for the crawler to visit in the next crawling step
$\{wp_i^{\text{good}}\}$	A set of "good" exemplary search result webpages manually identified by a human searcher
$\{wp_i^{\text{bad}}\}$	A set of "bad" exemplary search result webpages manually identified by a human searcher