(2) *What important information in the full text does not appear in the abstract?*[7]

(3) *What should an ideal summary of the full text contain that is not already in the abstract?*

(4) *What are the differences in the way authors and peers see an article?*

We explore these questions indirectly, using an under-explored information source: the sentences containing the citations to a target article or *citances*. While cocitation analysis is commonly-used for determining the popularity, and by association, the importance of a publication [8–15], our focus here is on the *contents* of the sentences containing the citations, that is, the citances.

In particular, we compare the information content of the abstract of a biomedical journal article to the information in all citances that cite that article, thus contrasting the important points about it as judged by its authors versus as seen by peer researchers over the years following its publication. Put another way, we use citances as an indirect way to access important information in the full text[8]. The idea is that (1) any information not mentioned in the abstract but referred to in citances should be coming from the full text, and (2) entities and concepts mentioned in a citance should be important and somewhat representative of their source.

To give an example, here is the abstract of an article (PubMed ID 11346650):

> *Multiple Mechanisms Regulate Subcellular Localization of Human CDC6.*
>
> *CDC6 is a protein essential for DNA replication, the expression and abundance of which are cell cycle-regulated in Saccharomyces cerevisiae. We have demonstrated previously that the subcellular localization of the human CDC6 homolog, HsCDC6, is cell cycle-dependent: nuclear during G(1) phase and cytoplasmic during S phase. Here we demonstrate that endogenous HsCDC6 is phosphorylated during the G(1)/S transition. The N-terminal region contains putative cyclin-dependent kinase phosphorylation sites adjoining nuclear localization sequences (NLSs) and a cyclin-docking motif, whereas the C-terminal region contains a nuclear export signal (NES). In addition, we show that the observed regulated subcellular localization depends on phosphorylation status, NLS, and NES. When the four putative substrate sites (serines 45, 54, 74, and 106) for cyclin-dependent kinases are mutated to alanines, the resulting HsCDC6A4 protein is localized predominantly to the nucleus. This localization depends upon two functional NLSs, because expression of HsCDC6 containing mutations in the two putative NLSs results in predominantly cytoplasmic distribution. Furthermore, mutation of the four serines to phosphate-mimicking aspartates (HsCDC6D4) results in strictly cytoplasmic localization. This cytoplasmic localization depends upon the C-terminal NES. Together these results demonstrate that HsCDC6 is phosphorylated at the G(1)/S phase of the cell cycle and that the phosphorylation status determines the subcellular localization.*

And here are some citances pointing to it:

> *Much of the soluble Cdc6 protein, however, is translocated from the nucleus to the cytoplasm when CDKs are activated in late G1 phase, thus preventing it from further interaction with replication origins [#C, #C and #TC].*
>
> *To ensure that the pre-RC will not re-form in S or G2, Cdc6p is phosphorylated and degraded in yeast (#C; #C; #C) or exported to the cytoplasm in higher organisms (#TC; #C; #C; #C; #C).*
>
> *It is phosphorylated by cyclin A-cdk2 at the G1-S transition and this modification causes some, but not all, of the Cdc6 to be exported out of the nucleus (#TC; #C; #C and #C).*
>
> *Cdc6CyΔ has a mutation in a cyclin binding motif that is an essential part of the substrate recognition signal for cdks (#TC).*
>
> *After entry into S phase, phosphorylation of HsCdc6, probably by cyclinA/CDK2, leads to its export from nucleus to the cytoplasm via NES [#TC].*
>
> *Once replication begins, Cdc6 is degraded in yeast (#C, #C, #C, #C, #C), whereas for mammals it has been suggested that Cdc6 is translocated out of the nucleus during S phase in a cyclin A-Cdk2- and phosphorylation-dependent manner (#C, #TC, #C,-#C, #C) and then subject to degradation by the anaphase-promoting complex (#C, #C, #C).*

In the above examples, #TC refers to the publication we are comparing against (the target citation: PubMed ID 11346650), whereas #C refers to other publications. Throughout this paper, we will refer to these citation sentences to other publications as *adjoining citations*.

Previous studies have discussed some of the potential of the use of *citances* for literature mining [16, 17]. Similar to anchor text on the web (visible, clickable text in a webpage, clicking on which navigates the user to another webpage), they are votes of confidence about the importance of a research article. Collectively, they also summarize the most important points about the target article, which makes them a potential surrogate for its full text [18] and an important knowledge source when generating a survey of scientific paradigms [19].

While previous work has focused on the *words* in citances, we compare their contents to the contents of the abstracts using coarse-grained biologically meaningful *concepts* such as entities, functions, and experimental methods.