The corpus of texts the KITAB project uses as the basis for its research is a subsection of the OpenITI corpus. It contains Arabic-language texts of the first 10 Islamic centuries – the OpenITI corpus also contains later texts, and texts in other languages.

One of the recurring questions we face is whether our corpus is representative of the Arabic written tradition. If statistical representativeness is meant (that is, are all characteristics of the whole Arabic written tradition present in the corpus to the same degree as in the whole tradition?) the answer obviously is that it isn’t, and can’t ever be. The corpus suffers from survival bias and selection bias: it only contains texts that somehow miraculously survived the combined effects of the natural enemies of paper (fire, water, light, vermin) and the fickleness of human interest. Moreover, of those texts surviving in manuscript form, only a small fraction has been printed and/or digitised; this selection has been driven by cultural, economical, religious and political factors.

That said, the question of what a mediaeval reader would think about the selection of books in our corpus is intriguing...

One way we can approximate this is by comparing the authors and books in the corpus to those mentioned in (bio-)bibliographical works, lists of books (and sometimes, authors) produced by dedicated book-lovers over the centuries. Although these bio-bibliographical works are not representative of the entire written tradition either, as they suffer some of the same survival and selection biases as the corpus itself (in addition to a geographical bias due to the places their authors resided in and/or visited), they do provide us with valuable snapshots of the books circulating/known at a specific point in time and space. For this first case study, we will use the [earliest](http://dx.doi.org/10.1163/1573-3912_ei3_COM_24336) extant bio-bibliographical work in the Arabic tradition, Ibn al-Nadim’s *Fihrist*.

Ibn al-Nadim (d. around 380/990) was a *warraq* (“stationer” – a seller of writing implements and books) who lived and worked at the intellectual centre of the Abbasid empire, Baghdad.[[1]](#footnote-1)

As observed by [Devin Stewart](https://www.jstor.org/stable/30069526),[[2]](#footnote-2) “The basic building blocks of the Fihrist are book lists, which fall into two categories: lists of books by a single author, which dominate the work, and lists of books in a specific genre”. The bulk of the work thus consists of short biographies of authors known to Ibn al-Nadim, with a list of the works they had written and which he had seen himself or read about in other books. The work is divided into ten chapters (*maqala*), each focussed on the authors who had written on a specific subject / in a specific genre:[[3]](#footnote-3)

1. On languages and scripts, revealed scriptures and Qur’an-related writings
2. On the Arabic language and grammar
3. On history and genealogy, and on rulers and administrators who wrote books
4. On poets and poetry collectors
5. On Islamic theology (subdivided into sections on 5 “schools” of theology)
6. On Islamic law (again subdivided into sections on 8 “schools” of law)
7. On Greek philosophy and sciences
8. On folk stories, magic, dreams, cooking, etc.
9. On the literatures of non-Muslim communities (but not Christians and Jews)
10. On alchemy

In this series of blogs, we will attempt to assess to what extent the OpenITI corpus overlaps with the Fihrist and what this can tell us about the wider representativity of both for the written Arabic tradition. The [next instalment](https://kitab-project.org/OpenITI-vs-Fihrist-part-2/) of the blog will focus on the methodology used, and the [final instalment](https://kitab-project.org/OpenITI-vs-Fihrist-part-3/) will describe the results of this study.

1. For a useful recent overview of the scholarship on Ibn al-Nadim, see Devin Stewart (2014). “Editing the Fihrist of Ibn al-Nadīm”, *Journal of Abbasid Studies* 1, 159-205. [↑](#footnote-ref-1)
2. Devin Stewart (2007). “The Structure of the Fihrist: Ibn al-Nadim as Historian of Islamic Legal and Theological Schools”, *International Journal of Middle East Studies* 39(3), 369-387: 370. https://www.jstor.org/stable/30069526 [↑](#footnote-ref-2)
3. For more on the structure and organisational principles of the Fihrist, please refer to Stewart 2007 and Shawkat Toorawa (2010). “Proximity, Resemblance, Sidebars and Clusters: Ibn al-Nadīm’s Organizational Principles in Fihrist 3.3”, Oriens 38(1-2): 217-247. https://doi.org/10.1163/187783710X536725. [↑](#footnote-ref-3)