Preface

In the ever-expanding realm of data-driven exploration and decision making, the ability to harness the power of natural language to interact with databases has become an indispensable pursuit. This book, "Natural Language Interfaces to Databases," presents a comprehensive journey through the fascinating world of Natural Language Interfaces to Databases (NLIDBs).

Chapter 1 serves as a gateway to this exciting topic, providing an Overview of NLIDBs. We begin with an interactive session with ChatGPT, providing a glimpse of the remarkable progress in this area and setting the stage for our discussion in the following chapters. We also highlight some of the pertinent challenges in deploying NLIDBs in real world scenarios, which have driven the advancements we witness today.

In Chapter 2, we establish a foundation by exploring the steps in building an NLIDB, covering essential basics using examples. This chapter introduces the intricate anatomy of NLIDBs, shedding light on the essential components that enable communication between human language and databases. Readers will gain a holistic understanding of NLIDBs' underlying mechanisms.

Chapter 3, Data and Query Model, presents the key concepts in representing, querying, and processing structured data, as well as approaches for optimizing user queries. Our discussion covers various query languages, such as SQL and Relational Algebra, while highlighting their connections to first-order logic. By doing so, this chapter equips readers with sufficient knowledge of database systems, enabling them to navigate through the remaining content of the book with ease.

The process of transforming textual queries into structured form takes center stage in Chapter 4, Text to data. This chapter provides a brief overview of early relevant work on semantic parsing and meaning representation before covering cutting-edge advancements in-depth. Readers will gain a comprehensive understanding of key topics on text-to-data translation that empower NLIDBs to comprehend and interpret human languages.

Chapter 5 covers the essential topic of evaluation, where we present various evaluation methodologies, metrics, datasets and benchmarks that play a pivotal role in assessing the effectiveness of mapping natural language queries to formal queries in a database and the overall performance of a system. Those metrics and benchmarks have been instrumental in advancing the field and providing us with the means to measure progress. By delving into the challenges of evaluating the equivalence between queries, readers will gain valuable perspectives on the continuous improvement of NLIDBs.

In Chapter 6, we explore the topic of Data-to-text, where formal representations of data are transformed to running text in natural language. NLIDBs, equipped with structured data, can generate human-readable narratives, presenting results in a coherent and contextually relevant manner. Crafting narratives from data is a powerful tool with applications in various domains. This chapter comprehensively reviews domain-specific

and domain-independent tasks that have been developed to evaluate the progress and improve the models in development.

Chapter 7 discusses Interactivity, the final frontier of NLIDBs. We introduce different dimensions of interactivity and the corresponding techniques for each dimension. We also discuss conversational NLIDBs and multi-modal NLIDBs where user input is beyond natural language. This chapter allows readers to have a comprehensive understanding of opportunities and challenges related to interactivity in NLIDBs.

Throughout the pages of this book, our endeavor has been to put together a balanced mixture of theoretical insights, practical knowledge, and real-world applications. Our earnest hope is that this compilation emerges as an invaluable resource, serving the interests of researchers, practitioners, and students eager to explore the fundamental concepts of NLIDBs.

In loving memory of Drago Radev, it is with a heavy heart and deep reverence that we present the final version of the book "Natural Language Interfaces to Databases." Our journey in crafting this work together spanned over more than two years, during which Drago's brilliance, unwavering dedication, and profound insights shaped every facet of its creation. His untimely passing in March 2023 was a profound loss, not only to those who had the privilege of knowing him personally but also to the academic and research communities worldwide. This book stands as a testament to his intellectual prowess and unyielding passion for knowledge dissemination. Although the completion of this work without his physical presence is bittersweet, we take solace in the knowledge that his spirit and ideas live on within these pages. It is our fervent hope that "Natural Language Interfaces to Databases" will continue to inspire, educate, and contribute to the advancement of our shared fields. May his legacy endure, and may this work serve as a tribute to the remarkable researcher who co-authored it — Drago Radev.

Let us embark on this journey of exploration into Natural Language Interfaces to Databases, which has become even more exciting with the arise of Large Language Models.

Yunyao Li
Davood Rafiei
July 2023