



Natural Querying

The difficulty in formulating complicated database queries could be said to come from two sources: cumbersome query language syntax (usually better tailored to the storage rather than its user) and unavoidably complex database structure. We usually can't do too much about the latter for performance reasons, but the former could undoubtedly be improved on.

Various, more natural querying languages have been proposed in the recent years, perhaps the most notable of them being FactEngine's "controlled natural language". But even there one does not avoid the need to learn a new language and to sit down to think through the more involved queries.

```
INSERT INTO conditions(time, location, temperature, humidity)
VALUES (NOW(), 'office', 70.0, 50.0);

SELECT * FROM conditions ORDER BY time DESC LIMIT 100;

SELECT time_bucket('15 minutes', time) AS fifteen_min,
location, COUNT(*),
MAX(temperature) AS max_temp,
MAX(humidity) AS max_hum
FROM conditions
WHERE time > NOW() - interval '3 hours'
GROUP BY fifteen_min, location
ORDER BY fifteen_min DESC, max_temp DESC;
```

```
WHICH Lecturer occupies WHICH Room
AND holds WHICH Position
AND is in WHICH School
AND is in A School WHICH is in (Faculty:'IT')
AND THAT Lecturer works for THAT Faculty
```

Goal

To develop an intuitive system for iterative database querying with natural language. Such a system needs to be able to understand the natural language in the context of query building and make use of the database schema and data stored to make informed guesses when faced with limited input.

Who is this for? Bachelor's or master's students interested in databases or natural language processing (NLP). Familiarity with either probabilistic or neural approaches to NLP is a big plus.

Interested? Please reach out to us for more details.

Point of contact: Peter Belcak, email pbelcak@ethz.ch, ETZ G63.