

Appendix A

SQL Database Schema for Squeal

These are the definitions for the SQL database schema defining Squeal (Chapter 2).

```
CREATE TABLE valstring
  (value_id INT NOT NULL,
   vcvalue VARCHAR(255) DEFAULT NULL,
   textvalue TEXT DEFAULT NULL,
   CHECK ((vcvalue = NULL AND textvalue != NULL) OR
          (vcvalue != NULL AND textvalue = NULL)))
```

```
CREATE TABLE urls
  (url_id INT NOT NULL,
   value_id INT UNIQUE NOT NULL,
   variant INT DEFAULT 1,
   PRIMARY KEY (url_id, variant))
```

```
CREATE TABLE parse
  (url_value_id INT NOT NULL,
   component CHAR(5),
   value VARCHAR(255) NOT NULL,
   depth INT DEFAULT 0,
   CHECK (component IN ('host','port','path','ref')))
```

```
CREATE TABLE page
  (url_id INT NOT NULL,
   bytes INT DEFAULT NULL,
   contents TEXT DEFAULT NULL,
   md5 CHAR(32) DEFAULT NULL,
   valid CHAR,
   CHECK (valid IN ('y', 'n', 'b')))
```

```
CREATE TABLE tag
  (tag_id INT NOT NULL,
   name CHAR(15) NOT NULL,
   url_id INT NOT NULL,
   startOffset INT NOT NULL,
   endOffset INT NOT NULL,
   PRIMARY KEY (tag_id))
```

```

CREATE TABLE att
  (tag_id INT NOT NULL,
   name CHAR(15) NOT NULL,
   value_id INT NOT NULL)

CREATE TABLE header
  (url_id INT NOT NULL,
   struct BINARY(6) NOT NULL,
   offset INT NOT NULL,
   ord INT NOT NULL)

CREATE TABLE list
  (url_id INT NOT NULL,
   struct BINARY(6) NOT NULL,
   offset INT NOT NULL,
   ord INT NOT NULL,
   tstamp TIMESTAMP)

CREATE TABLE link
  (source_url_id INT,
   anchor_value_id INT,
   dest_url_id INT,
   hstruct BINARY(6),
   lstruct BINARY(6),
   compute_id INT,
   tstamp TIMESTAMP)

CREATE TABLE rcontains
  (url_id INT NOT NULL,
   value VARCHAR(255) NOT NULL,
   helper CHAR(16) NOT NULL)

CREATE TABLE rlink
  (source_url_id INT,
   anchor VARCHAR(255) DEFAULT NULL,
   dest_url_id INT DEFAULT NULL,
   helper CHAR(16),
   CHECK ((dest_url_id = NULL AND anchor != NULL) OR
          (anchor = NULL AND dest_url_id != NULL)))

```