Create the Database

Parses 'pubs.txt' to create an in-memory structure called 'records' (list of dictionaries), uses 'records' to insert rows into relations 'publication', 'author', and 'written_by', and then pickles 'records' to avoid parsing on successive runs.

If the database 'pubs.db' has already been created, notify the user and exit.

```
In [1]: run create.py 'pubs.txt'
```

```
Parsing records...these will be used to write to the database.

Executing parsePublications...
Elapsed time: 3.06453204155

Creating database tables and inserting records...

Executing connectToDB...
Elapsed time: 0.000131845474243

Executing createTables...
Elapsed time: 0.0154640674591

Executing insertRows...
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Elapsed time: 14.266862154

Total elapsed time: 20.8824698925

Number of records:
        publication: 157804
        author:       83513
        written_by:  624756
```

Executing Queries

```
In [2]: %load_ext sql
        %sql sqlite:///database.db
```

```
Out[2]: u'Connected: None@database.db'
```

(1) Insert a new publication into the database.

```
In [3]: %%sql
        insert into publication values(157804, "Big Data and Recommender Systems", 2016, "", "");

        insert into author values(83513, "David C. Anastasiu");
        insert into author values(83514, "Evangelia Christakopoulou");
        insert into author values(83515, "Shaden Smith");
        insert into author values(83516, "Mohit Sharma");

        insert into written_by values(157804, 83513);
        insert into written_by values(157804, 83514);
        insert into written_by values(157804, 83515);
        insert into written_by values(157804, 83516);
        insert into written_by values(157804, 12857);
```

```
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
1 rows affected.
```

```
Out[3]: []
```

(2) Query all publications by a particular author.

```sql
In [4]: %%sql
        select p.id as ID, p.title as Title, p.booktitle as Journal,
            p.year as Year, p.pages as Pages
        from publication as p, written_by as w, author as a
        where p.id = w.pub_id
        and w.author_id = a.id
        and a.name = "Geoffrey E. Hinton";
```

Done.

Out[4]:

ID	Title	Journal	Year	Pages
1553	SMEM Algorithm for Mixture Models	NIPS	1998	599-605
1562	SMEM Algorithm for Mixture Models	Neural Computation	2000	2109-2128
54820	Exponential Family Harmoniums with an Application to Information Retrieval	NIPS	2004	
142790	GEMINI: Gradient Estimation Through Matrix Inversion After Noise Injection	NIPS	1988	141-148

(3) Query all publications in a particular year.

```sql
In [5]: %%sql
        select id as ID, title as Title, booktitle as Journal,
            year as Year, pages as Pages
        from publication where year = 1947;
```

Done.

Out[5]:

ID	Title	Journal	Year	Pages
125012	A Note on Quine's Principles of Quantification	J. Symb. Log.	1947	130-132

(4) Query all publications published in a particular journal.

```sql
In [6]: %%sql
        select id as ID, title as Title, booktitle as Journal,
            year as Year, pages as Pages
        from publication where booktitle = "NIPS"
        limit 20;
```

Done.

Out[6]:

ID	Title	Journal	Year	Pages
339	A High Performance k-NN Classifier Using a Binary Correlation Matrix Memory	NIPS	1998	713-722
624	A Comparison of Dynamic Reposing and Tangent Distance for Drug Activity Prediction	NIPS	1993	216-223
625	A Framework for Multiple-Instance Learning	NIPS	1997	
781	Multiple Instance Learning for Computer Aided Diagnosis	NIPS	2006	425-432
1538	Active Learning For Identifying Function Threshold Boundaries	NIPS	2005	
1551	Parametric Embedding for Class Visualization	NIPS	2004	
1552	Deterministic Annealing Variant of the EM Algorithm	NIPS	1994	545-552
1553	SMEM Algorithm for Mixture Models	NIPS	1998	599-605
1554	Parametric Mixture Models for Multi-Labeled Text	NIPS	2002	721-728
1555	Application of Variational Bayesian Approach to Speech Recognition	NIPS	2002	1237-1244
2194	Learning Body Pose via Specialized Maps	NIPS	2001	1263-1270
4479	Advances in Neural Information Processing Systems 5, [NIPS Conference, Denver, Colorado, USA, November 30 - December 3, 1992	NIPS	1993	
4480	Robust Learning of Chaotic Attractors	NIPS	1999	879-885
4481	Effects of Noise on Convergence and Generalization in Recurrent Networks	NIPS	1994	649-656
4482	Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping	NIPS	2000	402-408
4483	Representation and Induction of Finite State Machines using Time-Delay Neural Networks	NIPS	1996	403-409
4484	Using Prior Knowledge in a {NNPDA} to Learn Context-Free Languages	NIPS	1992	65-72
4485	Encoding Geometric Invariances in Higher-Order Neural Networks	NIPS	1987	301-309
4486	Extracting and Learning an Unknown Grammar with Recurrent Neural Networks	NIPS	1991	317-324
4487	Higher Order Recurrent Networks and Grammatical Inference	NIPS	1989	380-387

(5) Query all publications with a particular string contained in the title.

In [7]:
```sql
%%sql
select id as ID, title as Title, booktitle as Journal,
    year as Year, pages as Pages
from publication where lower(title) like "% ai %" or lower(title) like "%artificial intelligence%"
limit 20;
```

Done.

Out[7]:

ID	Title	Journal	Year	Pages
2430	The Financial Crimes Enforcement Network AI System (FAIS) Identifying Potential Money Laundering from Reports of Large Cash Transactions	AI Magazine	1995	21-39
4394	Presenting and Analyzing the Results of AI Experiments: Data Averaging and Data Snooping	AAAI/IAAI	1997	362-367
6024	AI 2005: Advances in Artificial Intelligence, 18th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, December 5-9, 2005, Proceeding	Australian Conference on Artificial Intelligence	2005	
6148	The Workshop Program at the Nineteenth National Conference on Artificial Intelligence	AI Magazine	2005	103-108
6287	Editorial - AI in the Life Sciences	Artif. Intell. Rev.	2003	7-11
7725	Logic for Programming, Artificial Intelligence, and Reasoning, 8th International Conference, LPAR 2001, Havana, Cuba, December 3-7, 2001, Proceeding	LPAR	2001	
7726	Logic for Programming, Artificial Intelligence, and Reasoning, 9th International Conference, LPAR 2002, Tbilisi, Georgia, October 14-18, 2002, Proceeding	LPAR	2002	
7727	Logic for Programming, Artificial Intelligence, and Reasoning, 10th International Conference, LPAR 2003, Almaty, Kazakhstan, September 22-26, 2003, Proceeding	LPAR	2003	
7728	Logic for Programming, Artificial Intelligence, and Reasoning, 11th International Conference, LPAR 2004, Montevideo, Uruguay, March 14-18, 2005, Proceeding	LPAR	2005	
7729	Logic for Programming, Artificial Intelligence, and Reasoning, 12th International Conference, LPAR 2005, Montego Bay, Jamaica, December 2-6, 2005, Proceeding	LPAR	2005	
7730	Logic for Programming, Artificial Intelligence, and Reasoning, 13th International Conference, LPAR 2006, Phnom Penh, Cambodia, November 13-17, 2006, Proceeding	LPAR	2006	
7732	Logic for Programming, Artificial Intelligence, and Reasoning, 14th International Conference, LPAR 2007, Yerevan, Armenia, October 15-19, 2007, Proceeding	LPAR	2007	
8173	A Comparative Analysis of AI and Control Theory Approaches to Model-based Diagnosis	ECAI	2000	136-140
8182	Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives	IEEE Transactions on Systems, Man, and Cybernetics, Part B	2004	2163-2177
9121	A combination Scheme of Artificial Intelligence and Fuzzy Pattern Recognition in Medical Diagnosis	FLAI	1993	157-164
10822	Adding AI to Web Services	AMKM	2003	322-338
11126	UAI '00: Proceedings of the 16th Conference in Uncertainty in Artificial Intelligence, Stanford University, Stanford, California, USA, June 30 - July 3, 200	UAI	2000	
12663	Does Money Matter? An Artificial Intelligence Approach	JCIS	2006	
16647	Erratum for: A Study of Homogeneity in Relational Databases [Annals of Mathematics and Artificial Intelligence 33(2) (2001) 379-414]	Ann. Math. Artif. Intell.	2004	443-444
17699	Artificial Intelligence: Methodology, Systems, and Applications, 11th International Conference, AIMSA 2004, Varna, Bulgaria, September 2-4, 2004, Proceeding	AIMSA	2004	

(6) Query all publications with a combination of the author, year, journal, and title constraints.

```
In [8]: %%sql
        select a.name as Author, p.id as ID, p.title as Title, p.booktitle as Journal,
            p.year as Year, p.pages as Pages
        from publication as p, author as a, written_by as w
        where p.id = w.pub_id and w.author_id = a.id
        and a.name = "George Karypis" and lower(p.title) like "%recommender systems%"
        and p.year > 2001 and p.booktitle = "";
```

Done.

Out[8]:

Author	ID	Title	Journal	Year	Pages
George Karypis	157804	Big Data and Recommender Systems		2016	

(7) Delete a publication from a particular author, year, journal.

```
In [9]: %%sql
        delete from publication
        where id in (
            select p.id from publication as p, author as a, written_by as w
            where p.id = w.pub_id and w.author_id = a.id
            and p.booktitle = "NIPS" and year = 1998 and name = "Geoffrey E. Hinton"
        );
```

1 rows affected.

Out[9]: []

(8) Delete all records for a given author.

```
In [10]: %%sql
         delete from publication
         where id in (
             select p.id from publication as p, author as a, written_by as w
             where p.id = w.pub_id and w.author_id = a.id
             and name = "Geoffrey E. Hinton"
         );

         delete from author where name = "Geoffrey E. Hinton";
```

3 rows affected.
1 rows affected.

Out[10]: []

(9) Change the name of an author or journal.

```
In [11]: %%sql
         update author
         set name = "Donnie Darko"
         where name = "Darko Marinov";
```

1 rows affected.

Out[11]: []

(10) Change the title of an article or its publication year.

```
In [12]: %%sql
         update publication
         set title = "Minority Report"
         where title = "SECRETS: a secure real-time multimedia surveillance system";
```

1 rows affected.

Out[12]: []