A SYSTEM FOR DIGITIZING WISCONSIN CRASH LOCATION INFORMATION

By

Arup Dutta[(1)]
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Wisconsin
Madison, WI 53706
Telephone: (608) 265-8583, Fax: (608) 262-5199
Email: arup@cae.wisc.edu

Steven Parker, Ph.D.
IT Project Manager
Traffic Operations and Safety (TOPS) Laboratory
Department of Civil and Environmental Engineering
University of Wisconsin
Madison, WI 53706
Phone: (608) 262-2591, Fax: (608) 262-5199
Email: sparker@engr.wisc.edu

Xiao Qin, Ph.D., P.E.
Associate Researcher
Traffic Operations and Safety (TOPS) Laboratory
Department of Civil and Environmental Engineering
University of Wisconsin
Madison, WI 53706
Phone: (608) 262-3649, Fax: (608) 262-5199
E-mail: xqin@engr.wisc.edu

Zhijun Qiu
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Wisconsin
Madison, WI 53706
Telephone: (608) 262-2524, Fax: (608) 262-5199
Email: qiuzhijun@wisc.edu

David A. Noyce, Ph.D., P.E.
Assistant Professor
Department of Civil and Environmental Engineering
University of Wisconsin-Madison
Madison, WI 53706
Telephone: (608) 265-1882, Fax: (608) 262-5199
Email: noyce@engr.wisc.edu

Prepared for the 86th Annual meeting of the Transportation Research Board,
Washington, D.C. January, 2007

Length of Paper:
6179 words+ 1 Tables and 6 Figures@250 words per table or figure, 7929 equivalent words

(1) Corresponding author

ABSTRACT

One of the key hurdles in identifying unsafe intersections and roadways in Wisconsin is the lack of a complete crash location map, especially for crashes that occurred on local streets. Crash locations are reported in terms of relative offset from an intersection, based on on- and at- street name information, which identifies the intersection, and direction and distance information, which identifies the offset. For intersection crashes, the offset distance is typically set to zero. This paper describes a system developed by the Traffic Operations and Safety Laboratory at the University of Wisconsin-Madison to automate the mapping of Wisconsin local road crash locations. The location mapping algorithm involves the integration of two separate Wisconsin Department of Transportation databases - the Wisconsin crash database of police traffic accident reports and the Wisconsin Information System of Local Roads (WISLR). The application of WISLR which includes an inventory of local roads such as traffic information, pavement condition and roadway geometry, provides invaluable access to more comprehensive safety analysis. Although the methodology introduced is specific to the two above-mentioned databases, the general ideas can be applied to any similar sets of crash and Geographic Information Systems (GIS) databases.

The final result is a pinpoint map of all the intersection and segment crashes that occurred on local roads in Wisconsin, along with the complete crash information associated with each mapped crash. The algorithm developed with this methodology is able to map approximately 79 percent of the intended pool of available crashes. Quality evaluations indicate that the mapping is almost 98 percent accurate.

Keywords: traffic safety, crash location identification, automatic map generation, Wisconsin Information System of Local Roads (WISLR)

BACKGROUND

During the ten year period from 1996 to 2005, there were on average over 130,000 police reported crashes annually on public roadways in the state of Wisconsin (*1*). Of those, nearly 5,500 crashes per year resulted in death or serious injury. It is clear that the need to mitigate crash hazards continues to be an important goal for Wisconsin transportation planners and engineers. Knowledge of both 'where' an incident occurs as well as 'how' the chain of events took place is necessary in order to improve the design and operations of intersections and minimize the consequences of a traffic accident. In general, however, there is no reliable system in place at this time to map all Wisconsin reported crashes onto a Geographic Information Systems (GIS) digital map for crash safety analysis.

In 1999, the Wisconsin Department of Transportation (WisDOT) Bureau of Traffic Safety (BOTS) secured federal Section 411 incentive grant funding to perform a Traffic Records Assessment that supported the development and activities of a State Traffic Records Coordinating Committee (TRCC). In 2002 and 2003, the TRCC gave top priority to automating crash records and improving location data. The Department of Motor Vehicles (DMV), during the same time period, successfully implemented the National Model Traffic and Criminal Software (TraCS) for internal entry of Wisconsin Motor Vehicle Accident Report (MV4000). DMV began pilot testing the new crash data collection system with four law enforcement agencies in September 2004.

DMV and BOTS have taken important steps to accelerate the deployment of TraCS, however the automation of crash data and location collection cannot take place in a short term because every police vehicle is not installed with a Global Positioning System (GPS) unit. Moreover, there is no plan in the near future to equip every vehicle in Wisconsin. As a result, it is clear that the process of automating crash location data collection and follow-up analysis is a long term prospect that will take several years to fully implement. Of additional concern is the fact that current deployment of the new technologies, software and recommendations to the MV4000 does not impact historical crash data; information which may be extremely valuable in the understanding of crash locations and patterns.

As a reaction to these needs, WisDOT has invested in the process of hand-coding all crashes that occur on state managed highways (STN) to a crash Reference Point (RP) and Relative Distance (RD) system for GIS mapping. However, there is still no ready way to map crashes that occurred on *local roads*. One solution - which is the one adopted by this project - is to develop an automated system.

Over the years, researchers have attempted to develop, analyze and disseminate crash related geocoding procedures and digital maps to suit the needs of their applications (*2-4*). Our study adds to the existing literature for developing methods to digitize historical crash data into geo-spatial maps. The purpose of this paper is to describe an algorithm and software tool developed by the Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin-Madison to automate the process of digitizing Wisconsin local roads crash information on a GIS map. In particular, the TOPS project had the following objectives:

- Develop an algorithm for automating the crash location data mapping process for crashes occurring in local roads with respect to existing WisDOT base maps and crash forms.

The algorithm will translate location information from a database of police crash reports to a geospatial map and create a pinpoint map from the crash information.

- Conduct a quality check for accuracy of the mapped crash locations for a sample subset of crash data.
- Determine the drawbacks, potential improvements and recommendations based on the results obtained.

DATA SOURCES

There are two primary data sources for this project: the WisDOT Crash Database of police reported crashes and the Wisconsin Information System for Local Roads (WISLR) GIS of geo-spatial information for all local roads in Wisconsin.

Wisconsin Crash Database

Wisconsin traffic crashes are, by statutory definition, "reportable" if someone is killed or injured, or if the property damage exceeds a certain threshold ($1000 for property related crashes or damage to government-owned vehicles and $200 for all other government-owned property, such as traffic control devices). Crash information is generally reported by a dispatched police officer via the Wisconsin MV4000 police form and is eventually archived in the WisDOT DMV crash database. The WisTransPortal data management system at the TOPS Laboratory contains a local copy of all crashes in the DMV database for the years 1994-2005. For the purposes of algorithm development, analysis, and quality evaluation, this project focused on crash records for the city of Madison, for the year 2003, downloaded from the WisTransportal. Figure 1 shows a subset of the records (rows) and fields (columns) from the WisTransPortal crash database. The definition for each field is provided below.

- *Accident-Number*: Computer system generated number to uniquely identify a crash. (This column has not been displayed in Figure 1 due to privacy concerns.)
- *On-Street*: The local street name on which a crash took place.
- *At-Street*: Name of the street which intersects with the street on which a crash took place.
- *At-Highway*: Name of the intersecting or nearest highway on which a crash took place.
- *Refpoint-Number*: State Trunk Highway reference point number where a crash occurred.
- *Accident-Location*: The type of location at which a crash occurs (public road intersection, public road non-intersection, parking lot, private property).
- *Intersection-Direction*: The direction from the listed intersection.
- *Intersection-Distance*: The distance from listed intersection location in hundredths of a mile.
- *Municipality-Code*: Identifies the Wisconsin municipality in which a crash occurred
- *On-Highway*: Lists the name of the highway on which the crash occurred.

On-Street	At-Street	At-Highway	Refpoint-Number	Accident-Location	Intersection-Direction	Intersection-Distance	Municipality-Code	On-Highway
PARKING LOT	POST RD			PL		0	1373	
UNIVERSITY AVE	N RANDALL AVE			I		0	1373	
S PARK ST	W BADGER RD			I		0	1373	
ACEWOOD BLVD2005	GOLDFINCH			N	S	1	1373	
N MILLS	UNIVERSITY AVE			N	S	1	1373	
W JOHNSON ST	WISCONSIN AVE			N		0	1373	
GAMMON PL				N		0	1373	
SUDBURY WAY	LAMPLIGHTER WAY			N		0	1373	
MARQUETTE ST	HAUK ST			I	E	0	1373	
		12	244	N		0	1373	39
FREEPORT RD	VERONA FRONTAGE RD			N	W	1	1373	
PARKING LOT	JOHN NOLEN DR			PL		0	1373	
MINERAL POINT RD	S HIGH POINT RD			I		0	1373	
PARKING LOT	JOHN NOLEN DR			PL		0	1373	
GORHAM ST	WISCONSIN AVE			I	W	0	1373	
WHITNEY WAY		12	340	I		0	1373	
N BROOM ST	W MIFFLIN ST			N		0	1373	
E WASHINGTON AVE	BALDWIN ST		73	I	E	1	1373	151
S STOUGHTON RD	E BUCKEYE RD			N	S	2	1373	
PARKING LOT	EAST TOWNE MALL			PL		0	1373	
OHMEDA DR	FEMRITE DR			I		0	1373	
E WASHINGTON AVE	N STOUGHTON RD	51	77	I	E	2	1373	151
ODANA RD	S WHITNEY WAY			I	W	2	1373	
	WESTBOUND	12	244	N	S	0	1373	39
GREENWAY TRL	CARNWOOD RD			I	S	1	1373	

FIGURE 1 Sample crash information from the WisTransPortal crash database.

Wisconsin Information System for Local Roads (WISLR)

The WISLR is a GIS based software package developed and maintained by WisDOT that combines data for local roads in Wisconsin with interactive mapping functionality. The database associated with WISLR contains roadway and intersection specific information that is used to build the geospatial map. According to WisDOT (*5*) “With WISLR, users can produce maps that show the location of road-related data and see trends that might otherwise go unnoticed. For this reason alone, WISLR aids with organized and logical assessments about local road data. This is just one example of what WISLR can do — and there are many other benefits.” WISLR was chosen for this project since it is the official local roads GIS database in WisDOT and, in addition, it provides an opportunity to link important physical roadway characteristic information to the crash reports for safety engineering analysis.

Figure 2 displays a portion of the City of Madison local roads map clipped from WISLR. It shows the nodes (intersections) and links (segments connecting two intersections) that is used as the base map for locating crashes. Understanding the relational information behind this map is critical to the development of the mapping algorithm. Many tables exist within the WISLR database. Detailed information about relations, tables and fields relevant to the mapping algorithm will be provided in subsequent sections.

FIGURE 2 WISLR links and nodes for the City of Madison.

OVERVIEW OF THE CRASH MAPPING ALGORITHM

This section provides a general overview of the key ideas and methodologies that were developed to map local road intersection and segment crashes. Mapping methodologies for intersection crashes are described first, followed by a description of segment crash mapping methodologies. Intersection crashes are those crashes which occur at or very close to a roadway intersection. Segment crashes are those crashes which occur within the roadway link which can be determined by two adjacent roadway nodes.

Specific implementation details are described more fully in the final Project Report which will be available on the TOPS Laboratory website (*6*).

Mapping Methodologies for Intersection crashes

The details of the intersection mapping methodologies are described below.

Intersection crash mapping data sources

Three fields from the crash database that were described earlier are required to map a crash record at the intersection level, namely— *Municipality-Code*, *On-Street* and *At-Street*. We emphasize that the location information in these fields originates from the MV4000 form which is hand coded by a police officer at the crash scene and subsequently manually entered into the DMV database.

The important WISLR tables and their relationships that are relevant to the determination of an intersection at or near which a crash occurs are shown in Figure 3. A description of each table is given below:

- **Roadway Route**: Unique list of road names for each municipality.
- **Alternate Roadway Route Prefix**: Contains standard and alternate prefixes for road names.
- **Alternate Roadway Route Name**: Contains standard and alternate spelling for common road names.
- **Alternate Roadway Route Type**: Contains standard and alternate types for road names.
- **Alternate Roadway Route Suffix**: Contains standard and alternate suffixes for road names.
- **Standard Roadway Route Prefix**: Contains standard prefixes for road names.
- **Standard Roadway Route Type**: Contains standard types for road names.
- **Standard Roadway Route Suffix**: Contains standard suffixes for road names.
- **On-At**: Contains the combination of road names that intersect each other.

The **Roadway Route** table contains a unique list of road names in WISLR separated into four parts: directional prefix, road name, road type and directional suffix. For example, "E Washington Ave" would be separated into three fields and an empty fourth field: *Roadway-Route-Prefix* = "E", *Roadway-Route-Name* = "Washington", *Roadway-Route-Type* = "Ave," and *Roadway-Route-Suffix* = "". Each street name in this table is associated with a unique *Roadway-Route-ID*. The four **Alternate** tables are used to take alternate spellings (aliases) that are part of a road name and standardize them. The three **Standard** tables contain the entire list of standard prefixes, types and suffixes used in WISLR. The role of the **Standard** tables will be described in subsequent sections.

In WISLR, intersections are identified with nodes listed in the **On-At** table. Each intersection is identified with a unique *Reference-Site-ID*. Two street names represented by their *Roadway-Route-ID* (the *On-Roadway-Route-ID* and the *At-Roadway-Route-ID*) form a node. We note that the **Roadway Route** table is represented twice in Figure 3 to illustrate the fact that a combination of the *On-Roadway-Route-ID* and the *At-Roadway-Route-ID* is required to obtain the *Reference-Site-ID*.

Intersection crash mapping methodology

The key to determining the intersection location for a crash record is to associate a WISLR *Reference-Site-ID* value in the **On-At** table shown in Figure 3 to a pair of *On-Street* and *At-Street* street names from the crash database.

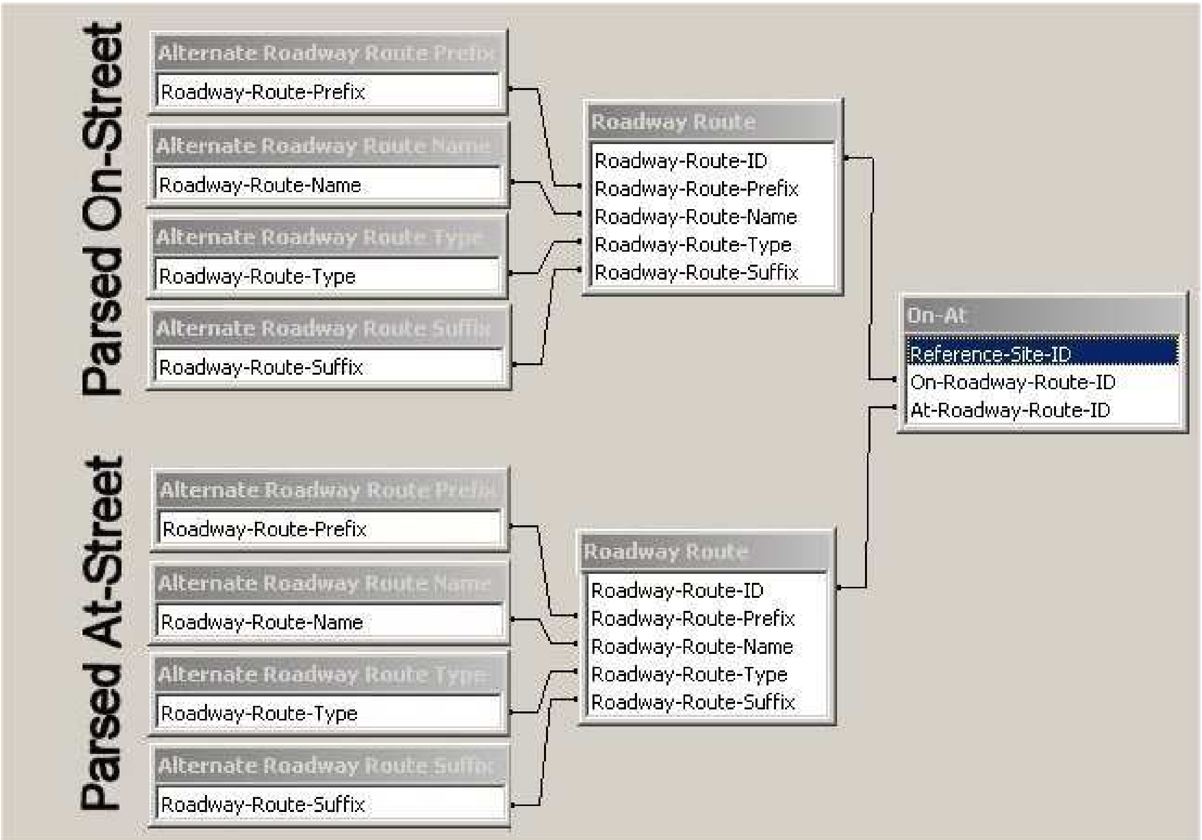

FIGURE 3 WISLR tables and relationships.

To create this association, the following steps are performed:

1. For each crash record, parse the *On-Street* and *At-Street* street names into prefix, name, type, and suffix components.
2. Match the parsed component information with the *Roadway-Route-Prefix*, *Roadway-Route-Name*, *Roadway-Route-Type* and *Roadway-Route-Suffix* information in the WISLR **Roadway Route** table for the same municipality in which the crash occurred. This gives a WISLR *Roadway-Route-IDs* for each *On-Street* and *At-Street* street name in the original crash record. Let these be represented by *On-Roadway-Route-ID* and *At-Roadway-Route-ID*, respectively.
3. Determine the *Reference-Site-ID* in the WISLR **On-At** table based on the *On-Roadway-Route-ID* and *At-Roadway-Route-ID*.

Once the *Reference-Site-ID* is obtained for a particular crash record, it is possible to generate an intersection level mapping since the *Reference-Site-ID* represents the intersection in WISLR where the crash occurred.

Implementation of the intersection crash mapping algorithm

The intersection crash mapping algorithm was implemented as a Java program to automate the process of assigning WISLR *Reference-Site-ID's* to crash database *Accident-Number's*. Specifically, the intersection mapping program performs the following steps:

1. Import crash information from a crash data file downloaded from the WisTransPortal.
2. Remove certain crash records from processing. This step is described further in subsequent sections.
3. Parse each crash record *On-Street* and *At-Street* field into its four parts: prefix, name, type, and suffix. Parsing is performed by splitting the information given in the *On-Street* (or *At-Street*) field into multiple words, and then utilizing the WISLR tables to analyze each word in order to determine if it should be a prefix, name, type or suffix. The following assumptions are used by the parsing mechanism:
    - i. At least one word in the parsed street name will be mapped onto the name field.
    - ii. Only the first word can be tested to see if it is a prefix.
    - iii. The last word can be tested to see if it is a suffix. If the last word is a suffix, the immediately preceding word can be tested to see if it is a type (if that does not violate Assumption i). If the last word is a not a suffix, it can be tested to see if it is a type.
    - iv. If a word is not a prefix, type or suffix, it has to be added to the name field.

The parsing mechanism performs two levels of analysis with respect to the WISLR tables. Level 1 analysis attempts to parse *On-Street* and *At-Street* fields into the prefix, name, type, and suffix fields based on the contents of WISLR **Standard** tables. Level 2 uses the **Alternate** tables in WISLR in order to convert non-standard formats into standards ones during the parsing procedure. For instance the alternate prefix 'North' could be standardized into 'N' in Level 2 parsing. Use of the **Alternate** tables is necessary since all street names are standardized in WISLR. We note that the algorithm only parses using **Alternate** tables if the street is not found in WISLR after parsing with **Standard** tables.

4. Match the parsed crash record street information with the contents of the **Roadway Route** table in order to obtain WISLR *Roadway-Route-ID's*: The match is done using a rigorous algorithm that considers spelling errors, roadway name aliases, and incomplete crash report information. The primary challenge in developing the matching algorithm was the existence of incomplete street name information in the crash records. To handle this situation, five levels of matching were established based on the amount of street name information used in the matching step:
    - i. Name Matching: The Name field of the parsed crash field is matched to the *Roadway-Route-Name* field in WISLR **Roadway Route** table. The additional Prefix, Type, and Suffix information is ignored.
    - ii. Prefix-Name Matching: Both Prefix and Name fields of the parsed crash record are matched to WISLR. Suffix and Type information are ignored.

iii. Name-Type Matching: Both Name and Type fields of the parsed crash record are matched to WISLR. Prefix and Suffix information are ignored.

iv. Prefix-Name-Type Matching. Prefix, Name, and Type fields of the parsed crash record are matched to WISLR. Suffix information is ignored.

v. Prefix-Name-Type-Suffix Matching: Prefix, Name, Type, and Suffix information of the parsed crash record are matched to WISLR This level takes into account all available street name information to find the WISLR *Roadway-Route-ID*.

Spelling errors are most critical in the Name field of the parsed crash record. If the Name field cannot be matched, all match levels would be unsuccessful. Therefore the Name field is assigned a 'spelling sensitivity' if all match levels are unsuccessful. A spelling sensitivity of *n* causes the matching process to compare only the first "*n* percent" of the characters in the crash record street name fields to the *Roadway-Route-Name* field in WISLR **Roadway Route** table. Spelling sensitivity can be implemented on the Name field while checking any of the five aforementioned match levels.

5. Match the *Roadway-Route-ID*'s to intersection *Reference-Site-ID's*: Given the *Roadway-Route-ID*'s, find the *Reference-Site-ID* which, as mentioned earlier, represents the intersection in WISLR.

After the description of the five steps, it is now possible to describe the intersection crash mapping algorithm that implements these steps. We first define the various parameters and their levels used in the algorithm:

i. parse_level= 1 or 2. parse_level=1 checks the parsed *On-Street* or *At-Street* against the **Standard** tables and parse_level=2 checks against the **Alternate** tables in WISLR.

ii. match_level= 1,2,3,4 or 5. Match level 5 performs a full prefix, name, type and suffix match whereas match level 1 performs only a name match.

iii. num_remove= 0 or 1. This parameter is introduced since it was found that some crash records had additional address information, i.e., street numbers included in the street name field for the crash record. Retaining this information would cause a mismatch in the name field. At the same time there may be streets present that have numeric names which are not street numbers (such as "1st St"). Therefore a parameter num_remove is introduced. num_remove=0 implies no numeric values were removed from street names. num_remove=1 implies all numeric values were removed from street names.

iv. spell_match=0 or 1. spell_match=0 implies that no spelling match is done for the name field. spell_match=1 implies that spelling match is performed for the name field.

The logic of the algorithm is to start with the most rigorous matching process and gradually relax the conditions until a successful match is found. In particular, the algorithm attempts to minimize any modifications of a street name given in the *On-Street* or *At-Street* field of the crash record (such as replacing possible aliases using **Alternate Roadway Route** tables, removing street address information, performing spelling match etc.) and at the same time attempting to find a match at the highest match_level value. If a match is not found, the match_level value is reduced. If no matches are found even at the lowest match level, then the least possible amount of modification of street name is introduced, and again there is an attempt to find matches from highest match_level to lowest match_level. This process is repeated until one or more *Roadway-Route-ID* is found or all match level and modification options are exhausted.

Segment crashes

Segment crashes are those crashes which occur within the roadway link which can be determined by two adjacent roadway nodes. The details are described below.

Segment crash mapping data source

Four fields from the crash database are required to map a crash record at the segment level, namely – *On-Street*, *At-Street*, *Intersection-Direction* and *Intersection-Distance*. The *Intersection-Direction* and *Intersection-Distance* fields provide information to locate roadway segment crashes relative to given intersections.

In order to map segment crashes, first the nearby intersection given by the *On-Street* and *At-Street* needs to be determined as described in the previous section. Next, two additional WISLR tables are required for mapping into a segment:

1. **Roadway Link**: Contains information for all roadway links. A roadway route can consist of several links. The field relevant to this discussion from this table is *Roadway-Link-ID*. The *Roadway-Link-ID* gives the identification in WISLR database for each individual roadway link

2. **Roadway Route Link**: Gives the relation between roadway links and roadway routes. At least one *Roadway-Route-ID* can be found in this table for each *Roadway-Link-ID*.

Segment crash mapping methodology

Segment crash mapping is essentially an extension of the intersection crash mapping algorithm. Segment crashes are coded in the crash database in terms of direction (*Intersection-Direction*) and distance (*Intersection-Distance*) from an intersection. Hence, the first step is to determine the unique intersection *Reference-Site-ID* related to each segment crash as described under intersection mapping. The key to determining the segment location for a crash record is then to determine the *Roadway-Link-ID* in the **Roadway Link** table in WISLR and shift the point from the intersection mapping result into a new location along the intersection direction based on the *Intersection-Direction* and *Intersection-Distance* information. There are several technical details, assumptions and procedures associated with segment crash mapping that are not covered in this discussion. Further details can be obtained from the final Project Report on the TOPS Laboratory website (*6*).

Interface development for intersection and segment crash mapping

The mapping procedure for intersection and segment crashes was implemented as a Visual Basic interface integrated with ESRI Map Objects 2.3. Specifically, the crash mapping program performs the following steps:

1. Import the list of *Reference-Site-ID*'s from the intersection crash mapping algorithm. The *Reference-Site-ID*'s are used to find all possible *roadway links s*urrounding the intersection, for each individual segment crash record.
2. Classify crashes as intersection or segment related using a customized algorithm that weighs several factors (*Accident-Location*, *Intersection-Direction*, and *Intersection-Distance*) from the crash record. The first step to identify whether a crash is intersection or segment related is based on the information given in the *Accident-Location* field ('I' representing intersection crashes and 'N' representing non-intersection, i.e. segment crashes, in this field). It is not possible however, in general, to reliably determine this directly from the *Accident-Location* field in the crash database, due to conflicting or insufficient information. For instance, some crash records are coded as segment related crash in the *Accident-Location* field, but the values of the distance to intersection are NULL. Other discrepancies and inconsistencies exist in the crash record, such as a value being assigned in the *Intersection-Direction* and *Intersection-Distance* field for crashes that are specified as intersection crashes. Based on discussions with WisDOT officials, a refined definition was developed for intersection and segment crashes. The refined definition is displayed in Table 1. The segment crash mapping algorithm only handles crash records that are categorized as segment related crashes in this refined definition.

TABLE 1 Refined definition of intersection related and segment related crashes

Accident-Location	*Intersection-Direction*	*Intersection-Distance*	I/S
I	NULL	[0, 2]	I
I	NULL	(2, +inf)	I
I	NOT NULL	[0, 2]	I
I	NOT NULL	(2, L)	S
I	NOT NULL	(L, +inf)	S
S	NULL	NULL	I
S	NULL	ANY	I
S	NOT NULL	[0,2]	I
S	NOT NULL	(2, L)	S
S	NOT NULL	[L,+inf)	S

Note: I/S is the refined identification of intersection related crash or segment related crash.

3. For segment crashes, use the *Reference-Site-ID* found in the intersection level mapping and *Intersection-Direction* from the crash record to identify the WISLR roadway link (*Roadway-Link-ID*) associated with the crash.
4. If a unique *Roadway-Link-ID* can be found, map to the segment using the Intersection Distance (*Intersection-Distance*) given in the crash report.
5. Generate a final crash map on the WISLR.

Figure 4 shows the Visual Basic interface that was developed for the automatic crash mapping program, named the Crash Mapping Automation Tool. The output of this application program is a shape (SHP) file which can be viewed by most of the popular GIS map software programs.

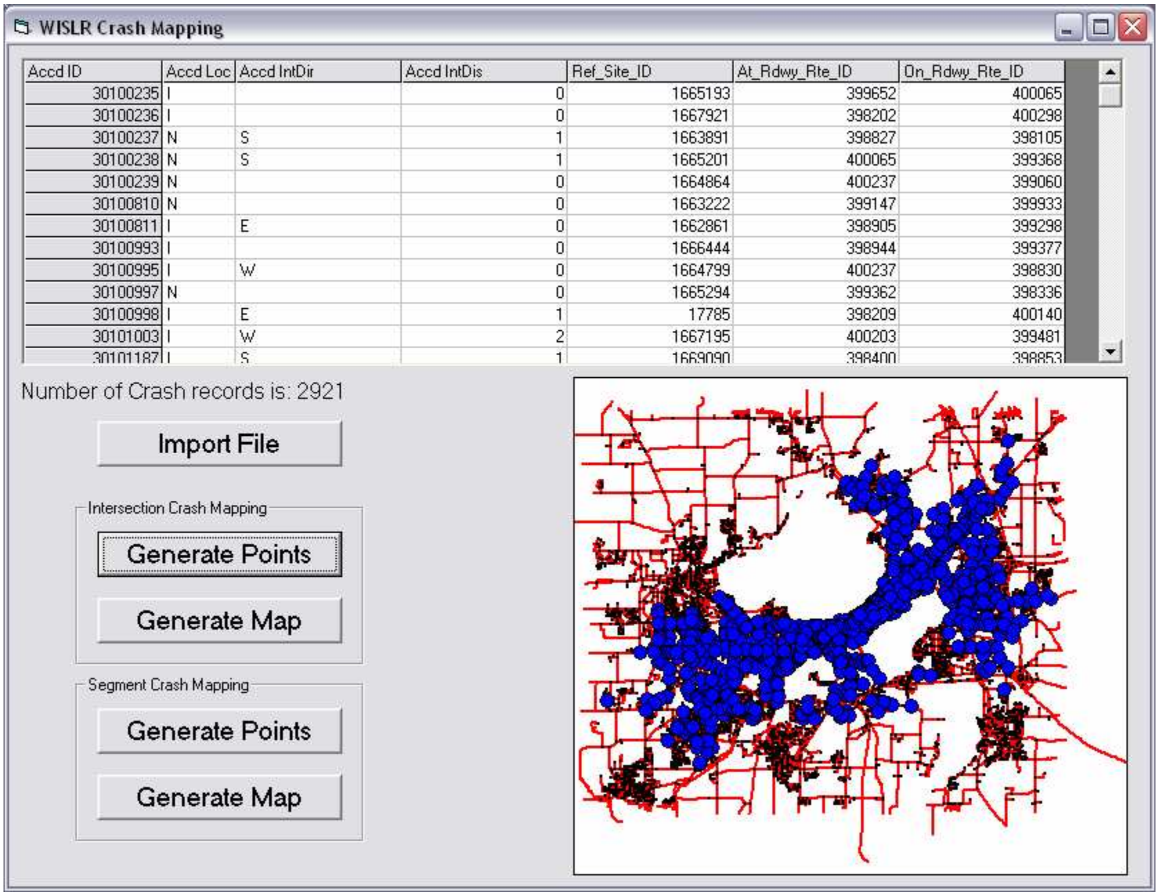

FIGURE 4 Interface for the WISLR crash mapping program.

CRASH MAPPING ALGORITHM RESULTS

This section describes the results of implementing the algorithms discussed in the previous section. The automated crash mapping process was tested in the developed software for crashes that occurred in the City of Madison in the year 2003.

Intersection and segment level mapping results

As described earlier, both intersection and non-intersection (segment) crashes are first mapped to the intersection level. Segment level mapping requires further processing once the intersection is found. This section presents results and analysis related to the intersection level crash mapping program. The results obtained through from the program are summarized below.

Filter results

At the start of the process, the crash mapping algorithm applies three filters to the raw crash data to exclude certain crash records from processing:

1. *Accident-Location* filter: Only public roadway crashes are processed. In particular, only crash records with Intersection (I) and Non-Intersection (N) *Accident-Location* types are retained by the mapping algorithm. Crashes marked as Parking Lot (PL) and Private Property (PP) crashes are excluded.

2. *On-Street* filter: An *On-Street* field in the crash record may be NULL (i.e. does not contain any information) for two reasons. First, the *On-Street* field information was not entered due to human error, so the data is incomplete. Second, it is possible that the street on which the accident took place was a highway, and is therefore not a local road accident. In either case, the records where *On-Street* field is NULL are removed from further processing.

3. *At-Street* and *At-Highway*: If both *At-Street* and *At-Highway* are NULL then the crash record has insufficient information to be mapped and is excluded from further processing. If *At-Street* is NULL but *At-Highway* is not NULL, then the record will be a candidate for mapping provided that *On-Street* is also not NULL.

A total of 5504 crashes occurred in the City of Madison in the year 2003. A total of 4351 crashes are available for mapping after applying the filters 1, 2 and 3.

Mapping results

The first step is to map both intersection and segment crashes to the intersection level. A total of 3,273 records were mapped to unique intersections; that is 75 percent of the 4,351 candidate records available for mapping after filter application and 59.5 percent of the 5,504 police reported crashes that occurred in the City of Madison in 2003. A total of 143 records were mapped to multiple locations (in such situations, the correct intersection is likely to be one of the multiple intersections found). Of these, 141 records were mapped to two intersections and 2 records were mapped to three intersections. Multiple mapped records generally occurred for two reasons. The first case is related to “horseshoe” structures on the local roadway system, i.e., where one road curves around and intersects a second road in two locations. In such a case, there is an ambiguity in the *On-Street* and *At-street* information. The second case is from incomplete street information in the crash table. Since the multiple records were mapped to only a few candidate locations (2 or 3 locations) it is believed that with some human intervention and judgment, they can be mapped quite easily to their correct location. Therefore, a total of 3,416 records mapped to intersections of the combined single and multiple mapped records; that is 78.5 percent of the 4,351 candidate records available for mapping after filter application and 62 percent of total 5,504 police reported crashes that occurred in the City of Madison in 2003. These 3,415 crashes consisted of both intersection as well as segment crashes, mapped to the level of the intersection. The remaining 936 crash could not be mapped due to several reasons, including spelling errors, incorrect/insufficient location information in the crash record, missing WISLR links and/or nodes etc.

Segment crashes were identified from these 3,416 crashes based on the refined definition given in Table 1. There were 594 crashes that were segment related, and all these crashes were found to belong to the sub-group of 3273 uniquely mapped crashes. In all, 590 of 594 segment related crashes can be mapped into a unique roadway link. Four of the segment crashes could not be adequately mapped in WISLR due to reasons such as the inability to get the correct directional information of the link, link errors in WISLR etc. Since such situations had rare occurrences, the details are not discussed here and can be obtained from the final Project Report on the TOPS Laboratory website (*6*).

In summary, of the 3,416 crashes that were first mapped at the intersection level, 2,679 crashes were identified and mapped as intersection crashes, 590 crashes were identified and mapped as segment level crashes, 4 crashes could not be mapped to segment level although they were identified as segment crashes, and 143 crashes were identified as intersection crashes but were mapped to multiple locations. For the 4 segment crashes that could not be mapped and the 143 multiple mapped crashes, these were all mapped to the intersection level. These 147 crashes would require some human intervention in order to be mapped more accurately.

Figures 5 and 6 present the mapped results for the 3273 uniquely mapped crashes up to intersection level. Figure 5 displays the frequency of crashes mapped up to the intersection level for all intersection and segment crashes. Figure 6 displays a sample of mapped segment crashes. Note that the segment crashes are offset from the intersection, as expected.

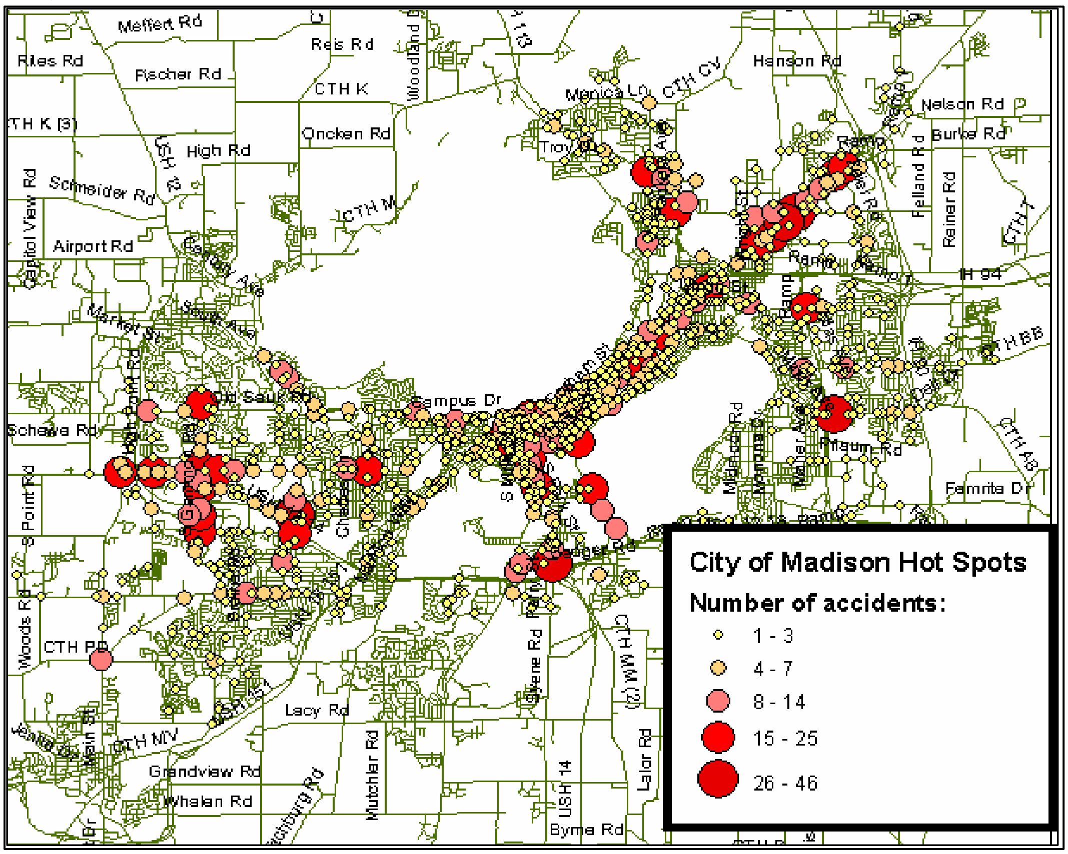

FIGURE 5 City of Madison (2003) hot spots.

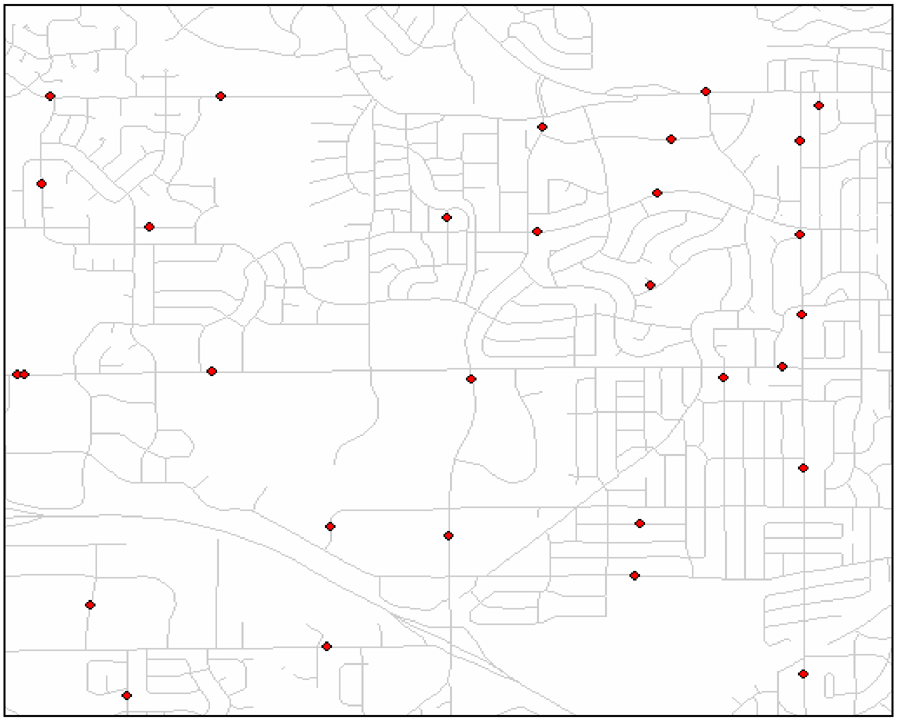

FIGURE 6 Sample of mapped segment crashes.

QUALITY CHECK OF MAPPING RESULTS

The crashes mapped by the algorithm were compared to manual mapping from two sources:

1. A digitized map of Madison 2003 crashes provided by the City of Madison. These crashes were digitized by hand directly from the MV4000 police reports.
2. A second comparison was performed by manually comparing the results for each crash with the Google Maps (*7*) online mapping service.

Quality evaluations were performed based on the comparisons for the crashes mapped through the developed algorithm. Additionally, the reasons for crashes that could not be mapped were also investigated.

Quality Check for Intersection Level Crashes

The City of Madison has developed a geo-spatial map in ArcGIS for all crashes that occurred in Madison in 2003, by manually locating each of the crash intersections. The geo-spatial map developed by the City of Madison was used to determine the accuracy of the crash locations given by the TOPS intersection level mapping algorithm. Due to inherent differences in the structure of the City of Madison map and the map generated by the algorithm – which is based on WISLR, only a limited number of crash records could be extracted for comparison purposes, and that too only for the crashes marked as intersection crashes. The locations for a pool of 1958 records for intersection crashes were extracted from this map and compared with the corresponding locations given by the intersection level mapping algorithm. Approximately 86 percent of the crash records matched in their locations and 2.6 percent were verified to be incorrectly mapped. For the remaining 11.5 percent of the records, the accuracy of the street locations were verified in WISLR for both the *On-Street* and *At-Street*, which means that it is quite likely that these were mapped to the correct intersection. However, a full verification for the 11.5 percent of the records could not be performed due difficulties in extracting their exact locations from the City of Madison developed map, and some shortcomings in the City of Madison developed map such as incorrectly located crashes.

One of the primary challenges in performing a quality check on the crash mapping results is that address information in the official police reports is not always clear. For a certain number of crashes, any manual mapping procedure will require a level of guesswork. As such, a second quality evaluation was performed by TOPS staff by comparing the results of the intersection level mapping algorithm to manual mapping using Google maps.

The 3,273 uniquely mapped records were manually mapped in Google Maps based on the *On-Street* and *At-Street* information in each record. The intersection identified by Google Maps for each record was then compared to the intersection to which the same record was mapped in WISLR. In particular, for each of the 3,273 records the location in WISLR map was first identified. Then the *On-Street* and *At-Street* information for the records were entered in Google Maps website (7) as: "<*On-Street* > & <*At-Street*>, Madison, WI".

When Google Maps was unable to locate exact matches for <*On-Street*> and/or < *At-Street*>, it provided alternate candidates for <*On-Street*> and <*At-Street*> names that could be interactively selected by the user. Under such a situation, the best candidate was selected based on the opinion of the evaluator. Then the crash location in Google Maps was visually compared to the crash location in the WISLR map given by the intersection level mapping algorithm. If the two locations were found to match, then the probability that the crash was mapped to the correct location was considered high.

Based on comparison of the intersection level mapping algorithm with manual mapping through Google, it was observed that manually mapped crashes on Google agreed with 98 percent of the 3273 records that were compared.

Quality Check for Segment Crashes

This section describes the results and analysis related to the implementation of the segment crash mapping algorithm. The algorithm was implemented for the same data source as the above intersection crash mapping, and results from intersection mapping are regarded as the input of segment crash mapping. In particular, the accuracy of segment mapping depends primarily on the accuracy of intersection crash mapping, as it is the precondition of finding the exact roadway link.

A minor factor of segment mapping errors is that the generated directional information associated with the roadway links was incorrect. This occurred in 4 of the segment crashes that could not be mapped. The primary reason for the incorrect directional information was that the roadway link in WISLR had too much curvature to match the directional information given by the crash record.

In summary, on the basis of our tested results over 99 percent of segment crashes obtained after intersection level mapping were accurately mapped to the correct roadway link.

Unmapped records

For records that were not mapped successfully, there were four primary reasons—spelling errors, missing WISLR intersections, alternate road names, and insufficient information in the crash records. Possible solutions to each of these causes are described below.

1. Spelling errors: The current spelling error handling module was able to map 329 records that would otherwise not have been mapped. In addition, there are 115 remaining crash records that could not be mapped due to spelling errors. Those records could potentially be mapped with sufficient improvements to the algorithm, which would represent a 2.5 percent improvement.

2. Missing WISLR intersections: For various reasons, some intersections are not included in WISLR. TOPS is working with WisDOT to track missing intersections by municipality for recommendation for inclusion in WISLR. Based on the analysis of the mapped dataset, a further 362 records could be mapped if such intersections are added, which represents an 8.3 percent improvement.

3. Alternate road names: A few known alternate names were identified manually, which resulted in the mapping of 29 additional records (0.6 percent improvement for the 4351 available records). However there were an additional 205 records that were mapped in Google Maps with manually entered alternate names known to the quality check evaluator, but did not have an intersection associated with them in WISLR. These 205 records did have the correct *On-Roadway-Route-ID* and *At-Roadway-Route-ID* for the alternate names in WISLR. If an alternate road table for each municipality could be created, and the missing WISLR intersections added, the combined improvement could map 596 additional records, i.e. a 13.7 percent improvement for the 4351 available.

4. Unintelligible records: A small percentage of records have either unintelligible or unmappable *On-Street* and/or *At-Street* fields (around 2 percent of the 4351 records). There is no simple solution to map these records. Any possible mapping would have to be manual based on guesswork.

The automated mapping algorithm provided an encouraging breakthrough in the ability to analyze local road crash information in Wisconsin. Further improvements are possible through a more advanced spelling error handling module, the development of alternate name roadway tables and the inclusion of missing intersections in WISLR.

NEXT STEPS

Quality assessment demonstrates that automatic crash location mapping yields high matching percentage and reliable outcome without sacrificing accuracy. The successful implementation of the mapping algorithm for 2003 City of Madison crashes provides great promise to safety analysis on local highways or streets. In order to maximize its benefits, future expansion of the tool to a statewide application is of immediate interest. One improvement is to make the tool portable so that the algorithm can be conveniently transferred to other municipalities or agencies that are in great need of the crash location information.

Knowing the crash location is the first step to identifying safety problematic locations. Providing versatility and flexibility is the key to the success of meeting users' needs. Consequently, the tool will be enhanced with more sophisticated query functions that allow users to conduct safety analysis on various bases such as hot-spot identification, corridor analysis or network screening.

Safety data is far more than just crash information. Conducting meaningful crash analysis and developing feasible safety improvements demand a variety of data sources. Using WISLR as the base map provides a seamless integration of crash information to highway inventory stored in the WISLR database. A user-friendly interface is required to facilitate more systematic and comprehensive safety analysis using valuable WISLR information such as roadway geometric characteristics, local traffic information, pavement conditions and intersection configuration, etc.

Ultimately, the tool can be upgraded to an Internet Interactive map using ESRI ArcIMS within the WisTransPortal framework. The upgrade not only provides maximum access to the highway safety community but also takes advantage of spatial query capability in ArcIMS such as proximity analysis and network analysis.

ACKNOWLEDGEMENTS

This research project was sponsored through the Federal Highway Administration (FHWA) state opportunity fund. The authors would like to thank the project Technical Advisory Committee members for their valuable comments: Jonathan DuChateau (WisDOT), Martha Florey (WisDOT), Susie Forde (WisDOT), and William Bremer (FHWA). Special thanks also go to Andrew Graettinger, Ph.D., for the technical advice and Jacob Notbohm for assisting in the quality assessment.

REFERENCES

1. Wisconsin Traffic Operations and Safety Laboratory, WisTransportal, University of Wisconsin, Madison. http://transportal.cee.wisc.edu/
2. Miaou, S., R. Tandon and J.J. Song. Providing Personalized Traffic Safety Information to the Public:Using Web-Based Geographical Information System Technologies. Report No. SWUTC/05/167424-1. *Texas Transportation Institute, 2005.*
3. Steiner, R., I. Bejleri, X. Yang, D. Kim, Improving geocoding of traffic crashes using a custom ArcGIS address matching application, *Proceedings of the 22nd Environmental Systems Research Institute International User Conference, San Diego California*, 2003, http://gis.esri.com/library/userconf/proc03/p0698.pdf
4. Zhan, C. Geocoding and analysis of freeway service patrol data. *Institute of Transportation Engineers Journal,* November, 2005.
5. WISLR website, Wisconsin Department of Transportation. http://www.dot.wisconsin.gov/localgov/wislr/index.htm
6. Traffic Operations and Research Laboratory, University of Wisconsin, Madison. http://www.topslab.wisc.edu
7. Google online mapping services. www.maps.google.com